AUGUST
1950
VOLUME 14
NUMBER 4

JOURNAL OF CONSULTING PSYCHOLOGY

AMERICAN PSYCHOLOGICAL ASSOCIATION

JOURNAL OF CONSULTING PSYCHOLOGY

EDITED BY LAURENCE F. SHAFFER
TEACHERS COLLEGE, COLUMBIA UNIVERSITY

ASSOCIATE EDITORS: EDGAR A. DOLL, *Devereux Schools, Devon, Pa.* WILLIAM A. HUNT, *Northwestern University.* E. LOWELL KELLY, *University of Michigan.* MORRIS KRUGMAN, *Board of Education, New York, N. Y.* BERTHA M. LUCKEY, *Board of Education, Cleveland, Ohio.* FRED MCKINNEY, *University of Missouri.* ANNE ROE, *New York, N. Y.* CARL R. ROGERS, *University of Chicago.* R. NEVITT SANFORD, *University of California.*

PUBLISHED BI-MONTHLY BY THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

THE JOURNAL OF CONSULTING PSYCHOLOGY is the clinical journal of the American Psychological Association. It is devoted primarily to the publication of original research relevant to psychological diagnosis, psychotherapy and counseling, personality, and the dynamics of behavior. Case studies, theoretical contributions, descriptions of clinical techniques, and discussions of the professional problems of clinical psychology also appear in its pages.

EDITORIAL OFFICE

Manuscripts submitted for publication, books and tests for review, and correspondence on editorial matters should be sent to the Editor, Laurance F. Shaffer, 525 West 120th St., New York 27, N. Y.

Information for contributors: The important characteristics of a good article are the significance of its subject matter, the demonstrated soundness of its conclusions, its clarity, and its brevity. Clarity is improved by careful outlining, discriminating use of headings, and a simple, straightforward style of writing. For brevity, authors should strive to convey every thought without waste of words, and should avoid repetitions of methods, data or findings.

Some of the mechanical requirements for manuscripts are: Type manuscripts with wide margins, on white bond paper. Double space between *each and every line*, including references, footnotes, matter to be set in smaller type, and tables. Use only two types of headings: centered headings, and underlined paragraph headings. Type each table, however small, on a separate sheet, with Arabic number and title. Draw all figures in India ink in final form for reproduction, with draftsman's lettering, but type captions for figures on separate sheets. The *References* must follow the exact form used in APA journals. For further information on the details of preparing manuscripts, see:

> ANDERSON, J. E., AND VALENTINE, W. L. The preparation of articles for publication in the journals of the American Psychological Association. *Psychol. Bull.*, 1944, 14, 345-376.

There is a lag of about nine to twelve months between the acceptance of an article and its publication. Authors may secure immediate publication, however, by arranging for articles to be printed as extra pages, increasing the number of pages that subscribers receive. The cost to the author of such early publication is approximately $11.00 per printed page.

Authors receive fifty reprints gratis, and may order additional quantities when returning proof. Authors are charged one-half of the extra cost of tabular material and cuts, and the full cost of changes made on the proof.

BUSINESS OFFICE

Communications concerning subscriptions, change of address, claims for the nonreceipt of a number, advertising, and other business matters should be sent to the American Psychological Association, Inc., 1515 Massachusetts Ave., N. W., Washington 5, D. C.

The Journal of Consulting Psychology is published bi-monthly in February, April, June, August, October, and December, the yearly volume comprising approximately 450 pages. The subscription per year is $5.00, foreign $5.50, single number $1.00.

Copyright 1950 by the American Psychological Association, Inc.

Printed and mailed by the Denman Printing Co., 23 West Colorado Ave., Colorado Springs, Colo. and entered as second-class matter at the Post Office at Colorado Springs, Colo. Acceptance for mailing at special rate of postage provided for in section 538, Act of February 25, 1925, authorized May 27, 1940.

JOURNAL OF CONSULTING PSYCHOLOGY

August, 1950 Vol. 14, No. 4

CONTENTS

The Concept of an Ideal Therapeutic Relationship: Fred E. Fiedler - - - - - - 239

Play Therapy with Well-Adjusted Retarded Readers: Robert E. Bills - - - - - 246

The Use of Limits in Reading Therapy: Muriel Potter - - - - - - - - - - - 250

A Client-Centered Approach to Vocational Counseling: G. D. Barahal, L. M. Brammer and E. L. Shostrom - - - - - - - - - - 256

A Factor Analysis of Rorschach Scoring Categories: J. R. Wittenborn - - - - - 261

Psychological Constriction on Several Projective Tests: Frances Merchant Carp - 268

Objective Rorschach "Signs" for Groups of Normal, Maladjusted and Neuropsychiatric Subjects: Roy M. Hamlin, George W. Albee, and Earl M. Leland - - - 276

Rorschach Human and Other Responses in Relation to Intelligence: John E. Tucker 283

Rorschach "Populars" as a Function of the Length of Record: Grace M. Thompson - 287

Discrimination of the Types of Schizophrenia by the Wechsler-Bellevue Scale: A. Edwin Harper, Jr. - 290

The Wechsler-Bellevue Psychometric Pattern in Anxiety Neurosis: Samuel J. Warner - 297

Intellectual Deterioration in the Aged: Agreement Between the Wechsler-Bellevue and The Babcock-Levy: Charlotte Fox and James E. Birren - - - - - - - - 305

Intelligence Test Scores of Newly Blinded Soldiers: Jacob Levine and Alan R. Blackburn - - - - - - - - - - - - - - - 311

Serial Versus Consecutive Order Administration of the Stanford-Binet Intelligence Scales: Arden N. Frandsen, Betsey R. McCullough and David R. Stone - 316

The Teaching of Psychotherapy Through Content-Free Interviews: Clare Wright Thompson and Katherine Bradway - - - - - - - - - 321

The MMPI as Diagnostic Differentiator: A Reply to Rubin: Bernard S. Aaronson and George S. Welsh - - - - - - - - - - - 324

A Note on "Reply to Rubin": Harold Rubin - - - - - - - - - - - - - - - 327

Correction of "The Relationship Between Variability and Ability on The Wechsler-Bellevue": Francis M. Gilhooly - - - - - - - - - - - 329

New Books and Tests - 330

THE CONCEPT OF AN IDEAL THERAPEUTIC RELATIONSHIP[1]

FRED E. FIEDLER[2]
THE UNIVERSITY OF CHICAGO AND VA MENTAL HYGIENE CLINIC

ONE of the major differences between the various systems of psychotherapy is the divergence of views concerning the most desirable therapeutic relationship that the therapist should create. There has been little, if any, disagreement among psychotherapists as to the paramount importance of a good therapeutic relationship to eventual cure [1, pp. 32-47; 2; 5, p. 3; 7; 8, pp. 85-114]. Rogers has been most insistent that the relationship is important; so have Thorne [10] and Horney [6, pp. 276-305]. Yet, the differences of opinion as to what constitutes a good therapeutic relationship are quite marked. Rogers favors a relationship which permits the patient most scope, and in which the therapist intervenes as little as possible, while Horney and Thorne advocate greater intervention and activity on the part of the therapist.

If, however, the therapeutic relationship itself is of such paramount importance then it seems unlikely that so many different types of relationships can be equally effective in helping patients to adjust, especially since all schools of therapy have reported successes as well as failures among their therapy cases.

One question we wish to answer here is, "Are these differences in theory semantic or do they represent actual divergencies in the goal therapists set for themselves?" If the theory and training of therapists are of greater importance than therapeutic skill, we will expect therapists of different schools to differ more from those of other schools than from therapists within their own school. If, on the other hand, experience and skill are of greater importance; that is, if the concept of what a therapeutic relationship ideally should be is derived from personal experience and skill, and only one such type of relationship has been found maximally effective, we will expect the more expert to agree among themselves irrespective of school, rather than with the less expert within their own school.

We can assume here that therapists will attempt to create the relationship which they consider ideal. We can further assume that the skilled therapist is better versed in his school's theory of therapeutic relationships than is the less expert. Consequently, the expert of any one school can be expected to disagree more with experts of other schools if the schools differ in theory, since experts are generally more representative of their school's practice and theory than novices of a school.

Stating this hypothesis in more operational terms, if there are any real differences in schools as to the relationship which they attempt to achieve we will expect a factor analysis to yield as many factors as there are points of view. If, on the other hand, these differences are present in theory only, and only one type of relationship is actually considered maximally effective, we will expect to find only one general factor among therapists of various schools.

Some therapists have expressed the conviction that the therapeutic relationship is something unique and does not occur outside of the therapeutic situation [i.e. 5 and 7]. We may assume therefore that one must have had experience with therapeutic situations in order to be able to describe them. An hypothesis would then be that lay persons, who have had no con-

[1]Reviewed by the Veterans Administration and published with the approval of the Chief Medical Director. The statements and conclusions published by the author are the result of his own study and do not necessarily reflect the opinion or policy of the Veterans Administration.

[2]The results of this paper were discussed in part at the Midwestern Psychological Association Meeting, Chicago, April, 1949. The writer wishes to acknowledge the suggestions he received from Mr. Ralph Heine and Dr. Morris I. Stein which were invaluable in clarifying some of the aspects of this investigation.

TABLE 1
PERSONAL DATA OF THERAPISTS AND UNTRAINED RATERS COOPERATING IN STUDY

Rater	Academic Training	Years Experience	Years Personal Therapy	Theoretical Orientation
PA-1*	AM	5	1	Psychoanalytic (Washington School)
PA-2*	MD	10	2½	Psychoanalytic (Chicago)
PA-3†	MD	10	7	Psychoanalytic
PA-4‡	MA, SSA	5	3½	Psychoanalytic social service
PA-5	AM	2	1½	Psychoanalytic
PA-6	MD	3	—	Psychoanalytic
ND-7*	PhD	6	1½	Nondirective
ND-8*	PhD	5	—	Nondirective
ND-9	AM	1	—	Nondirective
ND-10	AM	3	¼	Nondirective
IP-11*	MD	26	—	Adlerian
Ec-12*	AM	1	—	Eclectic
Ec-13	AM	5	—	Eclectic
Un-14	PhB	—	3‖	Untrained
Un-15	BS, RN	—	—	Untrained
Un-16	BS, EE	—	—	Untrained

*Considered to be very good therapist.
†This rater thought there "was a catch" to the ratings.
‡Experience of this rater has been primarily in social service work.
‖Psychoanalysis.

tact with psychotherapeutic situations, would be unable to describe the therapeutic relationship with any degree of success as compared to experienced therapists.

METHODS AND RESULTS

The method of investigation used here was to have therapists describe the relationship which they consider ideal. This was done by means of a series of Q-technique[3] [9] ratings by therapists of different schools and with different degrees of training and reputed skill. Two similar investigations were undertaken.

The First Study. The first utilized the ratings of six psychologists and two psychiatrists. Four of these therapists subscribed to a psychoanalytic theory, two were nondirectively oriented, and two considered themselves eclectics [See Table 1]. Among the four psychoanalytically oriented subjects, one had completed his medical and most of his psychoanalytic training, and had practiced for ten years. The other medically trained therapist had been practicing for three years and had not yet started his analysis. One psychologist had two years of intensive and five years of general therapeutic experience, and 150 hours of personal analysis. The other psychologist had little experience in doing therapy, but had received 120 hours of psychoanalysis.

Of the two nondirective therapists, one had had over 70 hours of personal nondirective therapy and had practiced for five years. The other had obtained 12 hours of nondirective therapy and had three years of experience. Neither of the eclectic therapists had received personal therapy. One had five, the other one year of experience as a therapist.

The measuring instrument was a series of 119 qualitative statements descriptive of patient-therapist relationships. These traits were randomly selected from a large number of statements about relationships which were drawn from books, journal articles, statements by therapists in case records, and in conferences. These 119 statements were mounted on cards, shuffled, and given to each subject with instructions to sort them into seven categories, with 2, 11, 28, 37, 11, and 2 statements in each category, with the statements most characteristic of their idea of an ideal therapeutic relationship at one extreme, and the statements

[3]Q-technique is a method developed by W. Stephenson which permits correlation between persons on the basis of traits. It is akin to inverse factor analysis.

least characteristic of the ideal therapeutic relationship at the other extreme of this normal distribution. Each of the statements, in accordance with the regular design of Stephenson's Q-technique, received a score from one to seven, depending on the category in which the card had been placed by the therapist. The rating by each therapist could then be intercorrelated with that by every other therapist, and the resulting correlation matrix factored by Holzinger's bi-factor method. The table of intercorrelations and factor loadings of each therapist are presented in Table 2.

TABLE 2

INTERCORRELATIONS* AND FACTOR LOADINGS OBTAINED IN THE FIRST STUDY

Raters	PA-1	PA-2	PA-5	PA-6	ND-7	ND-10	Ec-12	Ec-13
PA-1	—	66	72	61	77	59	70	64
PA-2		—	70	57	72	63	64	60
PA-5			—	61	78	67	69	68
PA-6				—	68	57	64	60
ND-7					—	64	72	67
ND-10						—	56	48
Ec-12							—	64
Ec-13								—
Factor Saturations:†	83	80	87	75	92	72	82	76

*Correlations were computed by the method of squared differences. All correlations are significantly different from zero on the 1 per cent level or above.

†$M_{resid.} = .002$; $SD_{resid.} = .027$.

The results indicate that all therapists correlate positively with each other. Correlations range from .48 to .78 with a median of .64. Reliability on resorting was .80.

The analyzed, psychoanalytically oriented therapists PA-1, PA-2, correlated with the nondirective therapist, who is considered an expert, (ND-7) .77, and .72, while they correlated with psychoanalytically oriented inexpert therapist, PA-6, .61, and .57. The differences between these correlations are significant at the 5 per cent level or above. The nondirective expert, who had correlated with the analyzed therapists .77 and .72, correlated only .64 with the less well trained nondirective therapist, ND-10, a difference also significant on the 5 per cent level or above.

The bi-factor analysis yielded only one general factor. Factor loadings for therapists who had received personal therapy are somewhat higher than loadings of others.

Inspection of individual items revealed no major differences between schools. Below are the statements which were characterized as most characteristic of an ideal therapeutic relationship.

An empathic relationship
Therapist and patient relate well
Therapist sticks closely to the patient's problems
The patient feels free to say what he likes
An atmosphere of mutual trust and confidence exists
Rapport is excellent
The patient assumes an active role
The therapist leaves patient free to make his own choices
The therapist accepts all feelings which the patient expresses as completely normal and understandable
A tolerant atmosphere exists
An understanding therapist
Patient feels most of the time that he is really understood
Therapist is really able to understand patient
The therapist really tries to understand the patient's feelings

The following statements were rated as least characteristic of an ideal therapeutic relationship:

A punitive therapist
Therapist makes the patient feel rejected
The therapist seems to have no respect for the patient
An impersonal, cold relationship
The therapist often puts the patient "in his place"
The therapist curries favor with the patient
The therapist tries to impress the patient with his skill or knowledge
The therapist treats the patient like a child

The Second Study. These results posed several questions: Are the results presented above a reflection of a stereotyped notion among therapists of what a good therapeutic relationship is? Are the results a function of the individuals who cooperated in this study, and the particular statements which were used here? Would an Adlerian expert also correlate highly with therapists of nondirective orientation? Finally, if, as some have suggested, the therapeutic relationship is something not found outside of therapy, will a lay person who had never been in therapy be unable to describe a therapeutic relationship?

A similar study was made to answer these questions. Seven different subjects cooperated (Table 1). A different array of 75 statements

about therapeutic relationships was selected on the following basis:

a. A number of therapists had to agree concerning the aspect of the relationship and the intensity of the statements, in order to make certain that statements had the same meaning to various therapists.

b. A number of therapists had to agree that all statements were descriptive of therapeutic relationships.

The relationship was envisiaged as consisting of three dimensions:

a. The therapist's ability to communicate with, and understand the patient.

b. The emotional distance which the therapist takes toward the patient (emotionally withdrawing, neutral, or close to patient).

c. The status of the therapist in relation to the patient (superior to the patient, subordinate to the patient, or equal with the patient).

To each of these dimensions were allotted 25 statements, and each dimension in turn was subdivided into groups of 5 statements each, representing five steps on each dimension. Statements were retained only when no disagreement could be found among therapists as to the dimension into which they should be placed, and the step within the dimension.

The instructions to the ten persons rating the statements were the same as in the previous study, except that statements were to be placed into categories with 1, 7, 18, 23, 18, 7, and 1 statements in each category.

The second study confirmed the results of the first study:

TABLE 3

INTERCORRELATIONS* AND FACTOR LOADINGS OBTAINED IN THE SECOND STUDY

Raters	PA-1	PA-3	PA-4	ND-7	ND-8	ND-9	IP-11	UN-14	UN-15	UN-16
PA-1	—	62	66	81	81	66	82	79	73	71
PA-3		—	43	50	61	44	62	60	57	58
PA-4			—	70	69	54	63	67	60	61
ND-7				—	82	69	76	78	73	72
ND-8					—	78	81	84	76	73
ND-9						—	63	72	65	56
IP-11							—	72	81	69
Un-14								—	79	65
Un-15									—	69
Un-16										—
Factor Saturations†	90	66	73	88	93	75	83	89	85	79

*All correlations are significantly different from zero on the 1 per cent level or above.
†$M_{resid.} = .003$; $SD_{resid.} = .033$.

1. Only one general factor was found; the same trends in correlations were apparent (See Table 3).

2. The Adlerian expert, IP-11, correlated highly with other well trained therapists.

3. The lay person who had been analyzed, Un-14, correlated .84 with one nondirective therapist, and .79 with a psychoanalytically oriented therapist. We are therefore justified in concluding that this concept of a good therapeutic relationship is not peculiar to therapists, especially since one person, who had never done, nor been in psychotherapy (Un-15) correlated .76 with a nondirective, .81 with an analytically oriented therapist, while she correlated only .56 with her husband, ND-9, who is a nondirective novice.

Inasmuch as only one general factor was obtained, the ratings could be pooled and a combined rating of the ideal could be obtained. This results from the pooling of the ratings of the four most expert therapists representing the psychoanalytic, the nondirective, and the Adlerian viewpoint. (PA-1, ND-7, ND-8, IP-11) Since their factor loadings are practically as high as the retest reliability of this second set of statements (.92), better estimates of the Ideal could not be obtained by pooling the ratings of more therapists.

This combined concept of the Ideal Therapeutic Relationship is presented below:[4]

[4]The numbering system of these statements gives the original classification into dimensions and steps within the dimensions:

Communication:

1-5 No communication is possible
11-15 Communication is poor
21-25 Some communication exists
31-35 Communication and understanding is good
41-45 Communication and understanding is excellent

Emotional distance

51-55 Therapist draws away from or rejects patient
61-65 Therapist is somewhat cool toward patient
71-75 Therapist is emotionally neutral
81-85 Therapist tends to draw emotionally close to patient
91-95 Therapist tends to be too close, is sticky

Status

101-105 Therapist feels very inferior and insecure
111-115 Therapist tends to look up and defer to patient
121-125 Therapist maintains peer relationship with patient
131-135 Therapist tends to look down on patient
141-145 Therapist feels very superior to patient

Category 7 — (Most characteristic of an Ideal Therapeutic Relationship:)

42. The therapist is able to participate completely in the patient's communication

Category 6 — (Very characteristic)

41. The therapist's comments are always right in line with what the patient is trying to convey
32. The therapist is well able to understand the patient's feelings
33. The therapist really tries to understand the patient's feelings
34. The therapist always follows the patient's line of thought
45. The therapist's tone of voice conveys the complete ability to share the patient's feelings
121. The therapist sees the patient as a co-worker on a common problem
123. The therapist treats the patient as an equal

Category 5 — (Somewhat characteristic)

21. The therapist reacts with some understanding of the patient's feelings
22. The therapist is able to keep up with the patient's communication much of the time
24. The therapist's reactions are neither particularly favorable or unfavorable in permitting free communication by the patient
25. The therapist usually maintains rapport with the patient
31. The therapist is usually able to get what the patient is trying to communicate
35. The therapist usually catches the patient's feelings
43. The therapist is never in any doubt about what the patient means
44. The therapist's remarks fit in just right with the patient's mood and content
71. The therapist is interested but emotionally uninvolved
72. The therapist's feelings do not seem to be swayed by the patient's remarks
73. The therapist maintains a friendly, neutral attitude throughout
75. The therapist shows little positive or negative emotion in his reaction to the patient
81. The therapist seems to like the patient
82. The therapist is pleasant to the patient
113. The therapist lets the patient determine the course of the session
122. The therapist gives and takes in the situations
124. The therapist acts neither superior nor submissive to the patient
125. The therapist treats the patient like a friend

Category 4 — (Middle category)

11. The therapist often flounders around before getting the patient's meaning
12. The therapist often misses the point the patient is trying to get across
15. The therapist's comments tend to divert the patient's trend of thought
23. The therapist's understanding of the patient's feelings is neither particularly good or bad
62. The therapist at times draws emotionally away from the patient
63. The therapist occasionally makes the patient angry
64. The therapist feels somewhat tense and on edge
65. The therapist seems to be a little afraid of the patient
74. The therapist accepts all of the patient's statements in a noncommittal manner
83. The therapist is pleased with the patient
84. The therapist is trying to establish an emotionally close relationship with the patient
85. The therapist is sympathetic with the patient
93. The therapist greatly encourages and reassures the patient
94. The therapist expresses great liking for the patient
95. The therapist is deeply moved by the patient
102. The therapist tries to sell himself
111. The therapist seems hesitant about asking questions
112. The therapist readily accedes to the patient's wishes
114. The therapist assumes an apologetic tone of voice when commenting
115. The therapist tries to please the patient
131. The therapist acts toward the patient in a somewhat protective manner
134. The therapist treats the patient like his pupil
135. The therapist directs and guides the patient

Category 3 — (Somewhat inapplicable)

3. The therapist somehow seems to miss the patient's meaning time and again
5. The therapist reacts in terms of his own problems
13. The therapist is unable to understand the patient on any but a purely intellectual level
14. The therapist finds it difficult to think along the patient's lines
53. The therapist is rejecting to the patient
61. The therapist is somewhat cool toward the patient
92. The therapist showers the patient with affection and sympathy
101. The therapist treats the patient like an honored guest
103. The therapist treats the patient with much deference
104. The therapist curries favor with the patient
105. The therapist always apologizes when making a remark
132. The therapist tends to look down on the patient
141. The therapist talks down to the patient as if he were a child
143. The therapist is very condescending to the patient
144. The therapist puts the patient in his place
145. The therapist gives the impression of feeling very much above the patient in social and intellectual status

Category 2 — (Uncharacteristic)

1. The therapist cannot maintain rapport with the patient
4. The therapist's own needs completely interfere with his understanding of the patient
51. The therapist feels disgusted by the patient
52. The therapist is hostile toward the patient
54. The therapist is punitive
55. The therapist is very unpleasant to the patient
142. The therapist acts in a very superior manner toward the patient

Category 1 — (Least characteristic of an Ideal Therapeutic Relationship)

2. The therapist shows no comprehension of the feelings the patient is trying to communicate

DISCUSSION OF THE RESULTS

In the two related studies which have here been presented, we have attempted to answer two questions:

1. Do theory and technique influence therapists' beliefs about the nature of the ideal therapeutic relationship?

2. Is the therapeutic relationship a unique phenomenon which exists only within the therapeutic situation?

Both studies have indicated that theoretical allegiance to one system of psychotherapy does not change the therapist's goal with respect to the relationship which he strives to create with his patient. The differences in factor loadings and correlations are not striking, and suggest that it is expertness which determines the type of relationship which is set as a goal by therapists. One would have expected at least some of the wide theoretical differences which exist, for instance, between nondirective and Adlerian therapeutic theories and techniques to be reflected in the assessments obtained from therapists of these schools. Yet, as in the first study, the two nondirective therapists, ND-7 and ND-8 correlate significantly higher (at the 5 per cent level or above) with the Adlerian therapist IP-11 than with the nonexpert of the nondirective school, ND-9. Similarly, they correlate more highly with the better trained psychoanalytically oriented therapist in this study than with the nonexpert of their own school. On the other hand, the psychoanalytically trained therapist PA-1 correlates more highly with the Adlerian and the nondirective experts than with the psychoanalytically oriented therapists PA-3 and PA-4, who are considered less expert. (The differences between these correlations are also significant at the 5 per cent level or above.)

The findings suggest that therapists generally agree on the most effective type of therapeutic relationship, and that theoretical differences are the result of poor communication between therapists of different schools.

It may be objected that the statements here used would naturally yield high correlations between persons since statements fall naturally into place. The fact yet remains that experts do have somewhat higher factor saturations, that only one general factor was found, and that significant differences were found between experts and nonexperts of the same school, as against correlations between experts of different schools. Finally, two different sets of statements were utilized, and a total of 16 different persons rated the statements. About ten other persons rated the ideal therapeutic relationship on the basis of these statements. Although these ratings were not incorporated into the factor design, the ratings thus obtained were thoroughly consistent with those here presented.

It may be objected that the statements do not really describe therapeutic relationships. Yet, a subsequent investigation which is now under way [4] definitely indicates that therapeutic relationships can be assessed by means of these statements by outside observers on the basis of wire recordings of therapeutic interviews.

The second question which this investigation posed concerned the uniqueness of the therapeutic relationship. Since the naive raters are practically indistinguishable from therapists on the basis of their ratings, it is difficult to see how such high agreement could be obtained if the therapeutic relationship actually were a unique phenomenon. These data rather support the hypothesis that a good therapeutic relationship is very much like any good interpersonal relationship. The latter is, no doubt, difficult to find in our society, yet one toward which our lay subjects can and do aspire.

The general factors which emerged from the correlation matrices do not seem to be associated with the therapist's theory, methods, or academic training. The only conclusion which we can draw from this is that all thera-

pists and naive subjects had essentially the same ideal in mind.

These studies leave us with several problems which remain to be solved:

1. Do therapists of the various schools actually create an essentially similar relationship with their patients?

2. Are the experts better able to approximate this generally held ideal than the less expert are capable of doing?

3. Is a good interpersonal relationship really like a good therapeutic relationship?

An investigation by the writer is now under way to answer the first two questions.

SUMMARY AND CONCLUSIONS

Two investigations of Q-technique design were undertaken to ascertain whether therapists with divergent theoretical views and therapeutic techniques also differed in their concept of an ideal therapeutic relationship. It was further attempted to answer the question whether the therapeutic relationship is unique to psychotherapy, or whether it resembles in general a good interpersonal relationship.

A number of therapists of different schools and with different degrees of training were asked to sort arrays of statements descriptive of therapeutic relationships into categories ranging from the most characteristic to the least characteristic. The ratings were then intercorrelated and the correlation matrices factor analyzed. In both studies only one general factor was found. The better trained therapists of different schools agreed more highly with each other than they agreed with less well trained therapists within their own school. Naive subjects were well able to describe the ideal therapeutic relationship.

The following conclusions could be drawn.

1. Therapists of different schools do not differ in describing their concept of an ideal therapeutic relationship.

2. The ability to describe this concept is probably a function of expertness rather than theoretical allegiance.

3. Nontherapists can describe the ideal therapeutic relationship in the same manner and about as well as therapists. The therapeutic relationship may therefore be but a variation of good interpersonal relationships in general.

4. A composite rating of the Ideal Therapeutic Relationship can be obtained by means of pooled ratings. This composite rating has here been presented.

Received September 20, 1949.

REFERENCES

1. ADLER, A. *Individual psychology*, translated by P. Radin. London: Kegan Paul, 1929.
2. ESTES, S. G. The therapeutic relationship in the dynamics of cure. *J. consult. Psychol.*, 1948, **12**, 76-81.
3. FIEDLER, F. E. Factors operative in shaping the concept of the ideal therapeutic relationship. Paper read at the Midwest Psychol. Ass., Chicago, April, 1949.
4. FIEDLER, F. E. A comparative investigation of early therapeutic relationships created by experts and non-experts of the psychoanalytic, non-directive, and Adlerian schools. Unpublished Ph.D. dissertation, Univ. of Chicago, 1949.
5. FREUD, S. *Vorlesungen zur Einfuehrung in die Psychoanalyse.* Vienna: Internationaler psychoan. Verlag, 1922.
6. HORNEY, KAREN, *Are you considering psychoanalysis?* New York: Norton, 1945.
7. MURRAY, J. M., AND BROSIN, H. W. General principles of psychotherapy. In H. L. Witmer (Ed.) *Teaching psychotherapeutic medicine.* New York: Commonwealth Fund, 1947. Pp. 273-288.
8. ROGERS, C. R. *Counseling and psychotherapy.* Boston: Houghton Mifflin, 1942.
9. STEPHENSON, W. Methodological considerations of Jungian typology. *J. ment. Sci.*, 1939, **85**, 185-205.
10. THORNE, F. C. Principles of directive counseling and therapy. *Amer. Psychologist*, 1948, **3**, 160-165.

PLAY THERAPY WITH WELL-ADJUSTED RETARDED READERS[1]

ROBERT E. BILLS
UNIVERSITY OF KENTUCKY

IN a recent article the writer [2] reported an experiment in which the reading ability of retarded readers significantly improved following nondirective play therapy. Seven of the eight subjects of the investigation showed severe emotional maladjustment, but the design of the investigation did not permit the conclusion that the treatment of the maladjustment was connected with the increased reading ability.

The present study was planned to test the hypothesis that significant increases in the reading ability of retarded readers will *not* occur when retarded readers who exhibit adequate emotional adjustments are given nondirective play therapy. If, as the result of experimental investigation, this hypothesis could be accepted, then the changes noted in the previous study [2] could be attributed to the treatment of the emotional maladjustment exhibited by the subjects.

DESIGN

The subjects of the experiment were eight well-adjusted third grade children from the University of Kentucky training school. These subjects were selected in the manner described below.

The experiment covered three periods of thirty school days each. On the first day of the first, or control, period the thirty children in the third grade class were given Form D of the Stanford Achievement Tests, Primary Battery, the California Test of Personality, and an incomplete sentences test[2] designed by the experimenter. During the next two weeks each child was administered the Revised Stanford-Binet, Form L. Reading retardation was defined as any negative discrepancy between reading age as measured by the Stanford Achievement Test and mental age as given by the intelligence test.[3] In this manner it was possible to isolate the children who were not reading at a level commensurate with their mental ability and to screen the adequacy of their emotional adjustments.

Twenty-one of the thirty children in the class showed some reading retardation. Of these twenty-one children, four were eliminated from consideration in the experiment because of the score obtained on the California Test of Personality, and four more were eliminated because of the incomplete sentences test. The remaining thirteen children were further examined by the Rorschach and the Thematic Apperception Test. Subjective evaluation of the Rorschach and TAT records was used to identify the eight retarded readers displaying the most adequate adjustments. On the basis of the test results these children were considered well adjusted.

The class was tested with Form E of the Stanford Achievement Tests on the first day of the second, or therapy, period and thus a measure of the expected gain of the experimental group in reading, spelling, and arithmetic during a period of thirty school days was obtained. During the second period each of the

[1]This investigation was made possible by a grant from the Research Fund Committee of the University of Kentucky. The writer deeply appreciates the assistance and cooperation of Lyman Ginger, Director of University School, and Miss Louise Willson, third grade teacher.

[2]This test was scored only for the direction of attitude toward self. The score was termed "per cent positive." This score was obtained by dividing the total of the positive and ambivalent attitudes by the total of the negative, positive, and ambivalent attitudes. A score of less than 50 per cent was considered to be sufficiently low to exclude a child from further consideration in the experiment.

[3]The writer appreciates the help of Lester Money, Elizabeth Ann Bicknell, and JoEllen MacCarter in the testing for the experiment.

eight selected children (four boys and four girls) was given individual and group play therapy following the principles of Rogers [3], Axline [1], and others. For each of the first three weeks of this period, each child received one forty-five minute individual therapy session and during each of the second three weeks an individual session plus a forty-five minute group therapy session. The boys and the girls attended separate therapy sessions. All individual sessions were recorded on wire and verbatim transcriptions were obtained from the recordings.

Form F of the Stanford Achievement Tests was given on the first day of the third period to measure the reading, spelling, and arithmetic gains of the class during the period when therapy was active.

Nothing was done with the children during the third period of the study, but at the end of this time they were examined for the fourth time with Form G of the Stanford Achievement Tests.

DATA

Table 1 contains the mental age, chronological age, intelligence quotient, reading age, and discrepancy between reading age and mental age of each of the eight children in the study. The data show that the children had an average IQ of 112 which does not differ much from the average IQ of the rest of the class, which was 110. The reading retardation of the experimental group varied from 6 months to 34 months; the average was 21 months.

TABLE 1

CHRONOLOGICAL AGE, MENTAL AGE, INTELLIGENCE QUOTIENT, READING AGE, AND READING RETARDATION OF THE CHILDREN

Name	CA	MA	IQ	RA*	Reading Retardation†
Lucy	8-11	11-4	127	10-1	1-3
Sally	8-6	9-10	116	8-1	1-9
Bill	9-0	10-6	117	9-5	1-1
Jack	8-8	9-4	108	7-2	2-2
Mary	8-10	10-0	113	7-11	2-1
Hank	8-7	9-4	109	7-2	2-2
Bob	9-10	11-2	114	8-4	2-10
Jane	9-0	8-4	93	7-10	0-6

*Reading age here is the average of the reading ages on the paragraph meaning test and the word meaning test of the Stanford Achievement Test.

†Reading Retardation is obtained by subtracting RA from MA.

The reading grade for each of the experimental children on each of the four tests and the gain (or loss) for each period of the experiment are given in Table 2. Tables 3 and 4 give the same data for the spelling and arithmetic tests. The data of Tables 2, 3, and 4 show that the gains in reading, spelling, and arithmetic do not differ markedly in the three

TABLE 2

THE READING GRADE FOR EACH CHILD ON EACH OF THE FOUR TESTS OF READING AND THE GAIN OR LOSS FOR EACH PERIOD OF THE STUDY*

Name	1st test	2nd test	3rd test	4th test	2-1	3-2	4-3
Lucy	5.05	5.45	5.70	5.25	.40	.25	—.45
Sally	3.05	3.65	4.15	4.00	.60	.50	—.15
Bill	4.45	5.15	5.20	5.10	.70	.05	—.10
Jack	3.20	3.40	4.00	3.95	.20	.60	—.05
Mary	2.95	3.25	3.40	3.95	.30	.15	.55
Hank	2.20	2.55	2.70	3.05	.35	.15	.35
Bob	3.30	3.85	3.75	3.40	.55	—.10	—.35
Jane	2.80	3.15	3.25	3.15	.35	.10	—.10

*The reading grade score given here is determined by averaging the scores on the tests of paragraph meaning and word meaning of the Stanford Achievement Tests.

TABLE 3

THE SPELLING GRADE FOR EACH CHILD ON EACH OF THE FOUR TESTS OF SPELLING AND THE GAIN OR LOSS FOR EACH PERIOD OF THE STUDY

Name	1st test	2nd test	3rd test	4th test	2-1	3-2	4-3
Lucy	3.8	4.3	4.7	4.5	.5	.4	—.2
Sally	2.4	2.5	2.7	3.4	.1	.2	.7
Bill	4.0	5.0	4.7	5.0	1.0	—.3	.3
Jack	2.6	3.0	3.0	3.8	.4	.0	.8
Mary	2.2	2.4	2.9	2.9	.2	.5	.0
Hank	2.3	2.0	3.0	3.1	—.3	1.0	.1
Bob	2.9	3.4	3.7	3.6	.5	.3	—.1
Jane	2.3	2.9	2.9	2.7	.6	.0	—.2

TABLE 4

THE ARITHMETIC GRADE FOR EACH CHILD ON EACH OF THE FOUR TESTS OF ARITHMETIC AND THE GAIN OR LOSS FOR EACH PERIOD OF THE STUDY*

Name	1st test	2nd test	3rd test	4th test	2-1	3-2	4-3
Lucy	4.00	4.00	4.30	4.30	.00	.30	.00
Sally	2.80	3.25	3.25	3.75	.45	.00	.50
Bill	3.75	3.95	4.10	5.05	.20	.15	.95
Jack	2.15	2.95	3.30	3.75	.80	.35	.45
Mary	2.85	3.10	2.80	3.55	.25	—.30	.75
Hank	2.50	2.50	2.75	3.30	.00	.25	.55
Bob	3.20	3.25	3.35	3.70	.05	.10	.35
Jane	2.95	3.00	3.10	2.95	.05	.10	—.15

*The arithmetic grade score given here is the average of the grade scores on the tests of reasoning and computation of the Stanford Achievement Tests.

periods of the study. The differences which are present may be accounted for by chance.

The recorded interview material and the data from the personality tests served to rate the children on their emotional adjustments. The writer and one of the examiners made a combined judgment of the adjustments of the children on the basis of the test and interview material. These judgments were made prior to examining the achievement test data. The rank for the adjustment of each child is given in Table 5 together with the rank for the total gain the child made in reading, spelling, and arithmetic. When these rankings are correlated a rho of .74 is obtained. This rho is not significantly different from zero since $N=8$ but with an n of this size statistical significance is a practical impossibility. Nevertheless, it is interesting that the rank of adjustment does agree so highly with the rank of total gain on the tests.

TABLE 5
RANKS OF ADJUSTMENT AND SUBJECT MATTER GAIN FOR THE CHILDREN

Name	Rank of subject matter gain*	Adjustment Rank	D	D^2
Lucy	4	5	1	1
Sally	6	3	3	9
Bill	2	1	1	1
Jack	7	7	0	0
Mary	5	4	1	1
Hank	8	8	0	0
Bob	3	6	3	9
Jane	1	2	1	1
	rho = .74			$\Sigma D^2 = 22$

*The rank of subject matter gain is the rank when all gains in reading, spelling, and arithmetic are totaled. All ranks are from least to most.

When the rank orders for the children in the experimental group in reading, spelling, and arithmetic are compared the coefficients in Table 6 are obtained. Similar correlation coefficients for the remaining children of the class are given in Table 7. None of these correlation coefficients is significantly different from zero, nor could they be expected to be when N is so small.

TABLE 6
THE RANK ORDER CORRELATION COEFFICIENTS OBTAINED BY CORRELATING THE READING, SPELLING, AND ARITHMETIC GAINS OF THE EXPERIMENTAL CHILDREN

	Spelling	Reading
Arithmetic	.94	.54
Spelling	—	.57

TABLE 7
THE RANK ORDER CORRELATION COEFFICIENTS OBTAINED BY CORRELATING THE READING, SPELLING, AND ARITHMETIC GAINS OF THE CHILDREN IN THE CLASS WITH THE EXCEPTION OF THE EXPERIMENTAL CHILDREN

	Spelling	Reading
Arithmetic	—.37	.00
Spelling	—	.19

DISCUSSION

The data of this study show that emotional maladjustment is not present in all retarded readers. The test and interview material indicate that varying degrees of adequacy of emotional adjustment existed in the children included in the study but none of the children could be classified as maladjusted.

Since the well adjusted retarded readers of this study failed to make statistically significant gains in reading ability following nondirective play therapy, it must be concluded that the data offer no reason for rejecting our hypothesis.

In the previously published study [2] it was concluded that the maladjusted retarded readers gained significantly in reading ability following nondirective play therapy. Insofar as can be determined, the children of the present study differed from the children of the previous study only in respect to the quality of their emotional adjustments. Since the well adjusted retarded readers, failed to make significant gains in reading ability following therapy and since the maladjusted retarded readers did exhibit such gains, it may be concluded that the reading gains noted in the previous study followed a nondirective treatment of the maladjustment present in the children.

When the ranking, by two clinicians, of the emotional adjustments of the children and the total gain in reading, spelling, and arithmetic were correlated a rho of .74 was obtained. This coefficient seems to indicate that clinicians can judge the emotional adjustments of children with a high degree of accuracy when sufficient material is available. The high correlation also

seems to indicate that the amount of gain made by the child is directly proportional to the amount of maladjustment present in the child and in this respect this study supports the previous study.

It appears from the data in Tables 6 and 7, that the gains which the experimental subjects made were evenly distributed in the fields of reading, spelling, and arithmetic. For the rest of the children there seems to be no relationship between the gain in any one field and the gains in the other fields. The test results indicate that for the nonexperimental children in the classroom the gains in one subject matter field were often made at the expense of gains in the other fields. Therapy seems to operate to produce even rather than irregular gains.

SUMMARY

By means of intelligence, achievement, and personality tests, eight well adjusted retarded readers were selected from a third grade class. These children were given a nondirective play therapy experience of an individual and group nature. As a result of this study and a previously published study [2] it was concluded:

1. Significant increases in reading ability resulted from a nondirective play therapy treatment of the emotional maladjustments which some retarded readers exhibit.

2. The gains in subject matter ability are directly proportional to the amount of emotional maladjustment present in the child.

3. When a child gains from nondirective therapy his gain is well-rounded and not specific to any one subject matter field.

Received September 7, 1949.

REFERENCES

1. AXLINE, VIRGINIA M. *Play therapy.* Boston: Houghton Mifflin Co., 1947.
2. BILLS, R. E. Nondirective play therapy with retarded readers. *J. consult. Psychol.*, 1950, **14**, 140-149.
3. ROGERS, C. R. *Counseling and psychotherapy.* Boston: Houghton Mifflin Co., 1942.

THE USE OF LIMITS IN READING THERAPY

MURIEL POTTER
TEACHERS COLLEGE, COLUMBIA UNIVERSITY

SO far as is known, no culture, primitive or complex, has ever existed in which well-defined limits were not set for acceptable behavior. The absence of such limits would imply the absence of a culture. However, social scientists have only recently achieved sufficient objectivity to enable them to recognize how narrow are these limits in any particular culture. Only after these limits were recognized could attempts be made to extend them experimentally, as in psychotherapy.

In psychotherapy with children the relationship between therapist and child is carried on at least as much by expressive activities as by verbalization. Each therapist sets his limits for the child's behavior in accordance with his philosophy and with the necessities of the physical environment. Just how permissive the therapist should be, to what extent the child may attack his person or destroy valuable artifacts in the environment, is debated with great intensity. Selecting and restricting the material environment is discussed in detail by Axline [2]. Even in a situation the purpose of which is to provide every opportunity for expressive and "aggressive" play, it is recommended that the toys and plastic materials should be easily manipulated and to the greatest extent indestructible. In such "self-directive" therapy the child's responses are just as truly limited by the environment as they would be limited by verbal statements in a situation where some of the artifacts — the therapist's clock or note-taking materials or person — are tabu for the child. Milieu limits are intrinsic to psychotherapy with children, just as they are intrinsic to all other human behavior.

When the limits of the milieu are not invoked by the use of carefully planned playroom, and sometimes when they are, verbal limits may sooner or later become a necessity.

Bixler [3] says, "It may be that the use of limits on behavior in therapy is equally as important as acceptance of the attitudes which provoke behavior" and defines very clearly what the limits on behavior are.

Axline [2] heads her chapter on *Limitations* with the statement, "The therapist establishes only those limitations that are necessary to anchor therapy to the world of reality and to make the child aware of his responsibility in the relationship."

Taft [8] speaks of the effective therapist as "someone who behaves as no one has ever behaved to him [the child] before: someone who understands and permits a use of herself, which determines for the client a new experience valuable, if at all, in and for itself."

Allen [1] defines these limits somewhat more clearly:

> But the realities of this experience soon bring about fresh sources of turmoil. Certain conditions define this relation. The child has an established time during which he is with the therapist. . . . . There are limits to the way he can live out his feelings because the therapist will maintain his own integrity and define certain rights of his own. The child can feel angry but when he wants to destroy or attack he will find a person who can stand steady before that force and set bounds to his actions. . . . .

Here are implied a specific set of limits which operate not only in therapy but in all interpersonal relationships:

1. The limits of the milieu. Behavior will be influenced by temporal and spatial factors—the length of the time available or allotted, the size of the therapist's office, the equipment he has for his own use, and the materials provided for the child's activities.

2. The limits of the program and process. It is the therapist who decides, though he may never explicitly need to state, what and how much (through the time limit) may or may not take place in the therapeutic interview. This is one of his contributions to the relation-

ship. It sometimes becomes explicit, as in a statement from Bixler's diary [3], "I had set several limits and finally had to tell him if he didn't stop throwing water on the floor I would have to send him out. He continued and I sent him out."

3. The limits set by the number of participants in the activity. In a two-person relationship the child is free to express his concerns, his ambivalence toward any object or problem, his resistance to the therapist (a) because no other people are present (b) because of the therapist's (limited) permissiveness, and (c) because what occurs between the two is not to be divulged. These limits might equally well, and paradoxically, be described as freedoms, but the child has these freedoms because the therapist always influences what happens in the relationship, whether by being directive, by being permissive, or even by refusing to participate. All social behavior is limiting and limited [7].

Recently a neurologist reported, after examining a nine-year-old nonreader, that his findings were negative; the youngster needed "reading and writing therapy." Reading therapy occupies a place midway between the treatment of behavior problems and the treatment of reading retardation by structured, preplanned lessons. Children who ask nothing better than the opportunity to receive special reading lessons do not require reading therapy. It is for the child referred to in the statement, ". . . .The teacher must break down the antagonism of the child who really resists instruction in reading" [4]. Briefly, its aims are to help the child, through his relationship with the therapist, (1) to develop a new attitude of optimism toward his reading difficulties; (2) to develop new attitudes toward all the people in the milieu in whose eyes he feels himself a failure—his teacher, his classmates, his parents, his siblings; (3) to use the emerging new attitudes in improving his reading skills in the most direct way acceptable to him.

In reading therapy, as in other kinds, limits are an essential part of the total dynamic situation. The points in the process at which either individual sets the limits, and the kind of limits set, are functions of the relationship. (When a child shows resistance to the therapist, he does so by setting limits! He doesn't like a book, he doesn't wish to read, he doesn't like the activities and passively refuses to participate.) Milieu and process limits are to be found, if sought for, in every secure and successful human relationship. They are utilized in good teaching and in cooperative group activities. In the treatment of reading difficulties they may appear either as techniques for reconditioning and reeducation, or as a dynamic aspect of the therapeutic process. They can be specifically described.

Limits of the Milieu. These limits operate at every moment in the reading therapy situation, and should be invoked from the beginning. An environment should be provided which sets *spatial* and *artifact limits,* and so controls the number of distractions to be resisted or responded to by the child. In the remedial reading room any materials are appropriate that are primary or accessory stimuli to reading activities: books and magazines on many subjects and of varied reading levels, art materials to be used for illustration and decoration of written work, tables and chairs of appropriate heights, comfortable for reading and writing, bulletin boards and blackboards, a typewriter and other printing materials, and so on. The purpose of this planned environment is to provide many stimuli to which the client may respond freely, without any limits having to be invoked, and as few as possible to which he may respond in a manner inappropriate to the aims of reading therapy. In such an environment it is possible for the therapist to be very permissive without losing sight of the program, which is activity relating to reading, however indirectly, or preparatory to reading, or actual reading. (However, an activity which in the therapist's judgment may be related to reading or language may not be so regarded by the child, and this relationship may not be made explicit by the therapist. Hence the "permissiveness" of the therapist may not be what it seems.)

Time limits have equal importance. If only a specified amount of time can be allotted to a child, this limitation is part of the reality which the child is meeting, in the same way that it operates in other kinds of individual therapy. The time limit, above and beyond administrative demands, must be set up with the needs of the individual child and his program

in mind, whether he is to receive 30 minutes of the therapist's time, or 45 minutes, or an hour, in accordance with the therapist's judgment of his ability to use it. The period may be used for a single activity, or it may of necessity be divided into a number of shorter periods occupied by different activities only a few of which are actual reading. The therapist is free to be permissive because the range of activities possible is small and related to the goals of the therapy. For example, conversation can be accepted and responded to freely because in a two-person situation the therapist, unlike the teacher, is free from the pressures attached to being responsible for a group of children, and can follow any direction in conversation or activity indicated by the child. The time limits are of course closely integrated with the therapist's formulation of the program, and with the interaction between client and therapist. A boy who for some days spent the first twenty minutes of his lesson time choosing a book finally met the limits by being told, "Perhaps you can't find a book because you don't feel like reading. If you don't find one by 2:30 (in ten minutes) let's not have a lesson—there won't be time to do any real work if we start later than that." Another youngster was given wider limits by the therapist, — "When you find a book you want to read I'll be glad to help you with it. Meanwhile I'll do my own work here at my desk until you are ready." Setting such specific limits is of course impossible until the relationship between therapist and child is a secure one.

A twelve-year-old boy who was unable to express the hostility he felt toward adults because of their control over him was for a long time unable to take advantage of the therapist's statement. "If you don't want to read any more, we can stop." Finally, on one occasion he was able to leave the room at the end of thirty-five minutes instead of remaining for fifty. Thereafter he no longer spent fifteen minutes at the beginning of the lesson time "deciding what to do."

Sometimes time limits can be utilized by the child to protect himself. "I'll read until 3 o'clock" (five minutes). A consistent length of time to which a "work" period can be held is helpful both for the number of different activities that can be carried on within it, and for the opportunity it gives the therapist to release the child if he shows fatigue or distaste. Such a release is part of the experience the child is obtaining in accepting responsibility for active participation in learning to read.

Limiting the number of meetings artificially is one kind of limit-setting not recommended. Usually the limits are set by reality—the length of a term, the beginning of a vacation, the demands made upon the therapist by other activities and plans. Reading therapy, like other kinds, cannot be hurried. The child must always be free to terminate the relationship; in that way he is a voluntary participant in it while it lasts. His voluntary participation should be based on his obtaining more satisfactions from its continuance than from its termination. It may be necessary at times to accept a child's resistance by terminating the lesson early — and sympathetically. At other times it may be advisable to verbalize the child's wish that the lesson were over without granting it. Each relationship will be different, and will require constant, active judgment on every occasion.

Program and Process Limits. The less a child accepts emotionally the fact that he needs reading lessons, the more latitude must be used in interpreting the activities he carries on in the lesson situation as appropriate. If he is ambivalent about participating in the learning situation, he will talk a good deal, often with great social adeptness, and read very little; or he will be excessively fatiguable or restless, or he will be unable to make any choice of the activities available. Frequently children who have been "good" and presented no school behavior problems show their feelings about reading disability by such signs. The more ambivalent the child, the more the reading therapist must recognize, reflect and interpret this fatigue, impatience, discouragement, reluctance, and even fear. "The print is too small," "You (the therapist) read this page, it hasn't any picture on it," "It will take too long to finish this," require sometimes reflection, sometimes reassurance, sometimes tacit support, sometimes a discussion with the child of his feelings and attitudes toward the situation. A yawn, a squirm or a facial expression may be a cue to the therapist's comment, "This gets tiresome after a while, doesn't it?", "Would you like to look out of the window for a minute to rest your eyes?", "Do you wish it was time to stop?"

"Shall we do something else now?"

Positive motivation and active participation are essential to learning reading skills, and conversely, passive participation and docile acquiescence by a child in a highly structured lesson sometimes produce peaceful lessons with but little retention. Since reluctance to participate appears, if not overtly, as extreme fatiguability or distractability or talkativeness, or as unwillingness to accept suggestions about activities, the therapist can help the child to be aware of his feelings and feel that they are shared, accepted and understood rather than condemned and rejected, by his reflective and interpretive comments. The child becomes freer to choose activities as soon as he knows that he may terminate them with equal freedom. Reading games sometimes give a youngster the opportunity to by-pass a particular reading activity without feeling guilty over having displeased or resisted the therapist, on whom the teacher's role is usually projected at first. However, participation in drill games is of no value until the child understands and feels that he has a need for the drill. Hence the limits the child sets by his choice or rejection of activities must be heeded and examined. The following situation is one in which program limits were invoked verbally when it seemed that the purpose of therapy was being denied by the child:

Twelve-year-old Bill, whose reading difficulty was symptomatic of a general emotional retardation, was very reluctant to settle down to specific reading activities. He had come to his first reading lesson unaware that plans had been made to this end. The therapist recognized his bewilderment and questioning him, found him unprepared. When asked if he wanted to have help with his reading, he replied honestly, "I do and I don't." Although frequently he was verbally positive about his attitudes toward lessons, his behavior expressed his unwillingness to participate in them. One device used by the therapist, on a day when he was particularly resistant, was a word bingo game [5]. The therapist in introducing it explained that it was "only reading with sugar on it," and Bill played it throughout the hour. At the following meeting the youngster began by taking a set of "Authors" cards out of his pocket and suggesting that the therapist play with him. The therapist's response was, for the first time, to set program and process limits in the following way: "No, we can't play a card game unless there is quite a lot of reading in it. Whatever you and I do here must always have some connection with reading, because you are coming for reading lessons." She offered to play bingo with Bill again, but he chose a book instead and settled down to reading without any further comment from her.

Children's reluctance to reread material and to participate in drill activities is based on sound premises. Providing vocabulary practice through reading several small books of the same reading level is more satisfying, because more interesting and success-providing, than rereading the same material. A satisfied, relaxed and optimistic child can learn more through two exposures to a word in interesting and challenging context than through ten successive exposures to which he passively submits. John, discovering that he was reading with about equal success two books one of which was at the second and one at the third grade level, was asked how this could be. He replied, "It's always easier to read something when it's interesting." Moreover, there are few books interesting enough to bear a second reading when one has as much reading experience to make up as retarded readers.

The more emotional concomitants of a reading difficulty, the more resistance is to be expected to structured teaching periods [6]. The length of time which a child uses in testing the limits is also prognostic of the rate of his progress. A nonreader spent an hour taking books from the shelves, turning the pages, and replacing them. Since she was unable to accept the therapist's offer to read aloud to her, she was judged unready for—that is, unable to accept—reading help. She could not participate in a program of reading activities no matter how little responsibility was placed on her—she could not even listen.

Language activities not involving reading may be of the greatest value, and well worth the time spent on them. Conversation which arises spontaneously out of the situation and the common experiences of child and therapist is an assurance to the child of his acceptance. Often it opens to the therapist new vistas for utilizing the child's interests; it gives opportunities for new insights into the child's attitudes and concerns. Eventually, however, limits may have to be set on the amount of nonreading language activities initiated by the child, which may be used deliberately or innocently as delaying tactics.

Before the relationship between the two is secure, suggestions by the therapist that the child return to reading activities when he seems to be avoiding them must be made neutrally, and refusals accepted with equanimity, but they should be made periodically. Frequently they will be accepted. Reflections and interpretations such as "I guess you don't feel much like reading today" must be made sympathetically if at all. The recognition of reluctance by asking a question such as "What would you be doing now if you hadn't come for a lesson?" may lead to a clarification of the child's ambivalence and a changed assignment of lesson time involving less sacrifice of the child's other interests. A few experiences of this kind remind the therapist that the child will be able to participate more freely in reading lessons at times when he will not be giving up school or extracurricular activities which he especially enjoys or which have prestige value for him. On the other hand, the amount of irrelevant conversation is also a measure of the amount of resistance. The therapist must set the limits by constant (but never verbal) evaluation of what the nonreading language activity contributes to the dynamics of the relationship or to program content.

Nine-year-old Timmy, who was using failure to read as a weapon against a family who expected him to grow up and meet all their demands without giving him any responsibility for himself in return, always preferred conversation to reading. For some time the therapist responded to and dealt with his extreme fatiguability, shown by yawning and restlessness, by suggesting changes in activities. After several months, when the relationship had been well established, the therapist could ask, "You would rather talk than read, wouldn't you?" The answer was a sidelong glance, then a grin. After this time the therapist would sometimes ask in a conversation, "How did we get so far away from reading?" as a means of returning to the task in hand. The day came when Timmy, in the midst of a conversation about wild animals in Africa, said to her, "Now I'm stalling!", laughed, and returned to his work.

At the point in the relationship where the child has become sufficiently secure to recognize his need for reading help by entering into brief and informal teacher-pupil activities, it is sometimes necessary that he be helped to succeed by gentle insistence and assistance in the completion of a reading task. At this point the therapist's teaching skill must extend to the provision of reading tasks of length and difficulty such that a small additional amount of effort and persistence results in measurable success. As a youngster reads one page more than last time, his satisfaction is increased by the feeling that he succeeded by his own effort. Thus it becomes easier and worth while to try again.

The term "increasing the level of tolerance" may well be used to describe this process. The need of the child who is ambivalent toward reading is to become slowly immunized against his intolerance of it, and this can be done only by small, easily assimilated, emotionally satisfying (success-providing) doses administered by an adult who is prepared to modify the schedule of dosage at any point where it shows signs of becoming an overdose. The therapist, to continue with this figure, must be prepared to change the amount and form of the reading treatment at the slightest sign of intolerance or inadequate assimilation. For children with severe reading difficulty reading has often become an experience so painful that opening a book or the very appearance of print produces a postural withdrawal response. In such cases the therapist must plan to give a wide choice of activities most of which may seem at first only very indirectly related to reading. The reading tolerance can be increased only by very cautious increases in the dosage.

Limits Set By the Two-Person Relationship. The experience of a child in the usual classroom situation is that of submitting to the classroom time schedule and to a number of rules and regulations which often seem arbitrary and meaningless to him. Class activities are group rather than individual; there is relatively little choice or freedom in either process or content. Since a large group must work and play together in the classroom, expressive and muscular activity are necessarily limited, and conformity is essential both for group comfort and for teacher approval. In the reading therapy situation, however, the therapist, although also the teacher, does not play a teacher role. When there are only two people present in a room, they may move about freely, speak aloud to each other whenever they wish without disturbing others, change their activities whenever they feel so inclined. They are not distracting others by making use of this free-

dom, nor are they being distracted by the presence of others. Particularly, the child is not distracted by the presence of other children, to which he has become negatively conditioned because of his reading failure. Moreover, he is free to express his resistance by nonconforming behavior which in the classroom situation would distract the other children and perhaps threaten the teacher's prestige. When the therapist recognizes and permits him to express his negative attitudes toward reading and toward the adult in the teacher's role without experiencing retaliation or condemnation, he is able to develop new attitudes in which his fear and guilt no longer handicap him in learning.

A twelve-year-old boy expressed his feeling about the story he was reading aloud by reading in baby-talk. He had himself selected the story because he knew it was of a reading level low enough to insure him success—a partial acceptance of his responsibility — but he had no illusions about the degree of his retardation, and he was very angry at the injustice of it. The therapist made no comment. She might have said, "It makes you mad to have to read this easy stuff, doesn't it?" but actually, silence was greater acceptance than any comment.

Because only two people are present, the child may hesitate, stop to think, change his response, look back over the material, read aloud, ask questions or the reassurance of the therapist, guess without embarrassment and be prompted as often as he forgets—and it *is* often—without suffering any of the humiliations that are implicit in failure as a member of a class. He learns that his mistakes in the therapy situation are being turned to good account because they are the clues to the kinds of lessons and exercises he needs. In this specially planned situation he learns to respect himself as a person who, though he has failed in learning to read, is still worthy of the attention and friendship of an adult whom he values. At the same time he discovers that he is not hopelessly different from other children, but that he has the same power to learn.

SUMMARY

Reading therapy attempts to set the stage for new and successful experiences with reading and with relationship to an adult. The therapist functions first, by playing the role of an adult who is neither parent nor teacher; second, by providing opportunity for reading experiences in which no guilt attaches to failure or nonconforming behavior and in which success is probable; third, by setting behavioral and milieu limits which are easy for the child to accept because the satisfactions in accepting them outweigh the advantages of rejecting them. Most retarded readers express their resistance and ambivalence toward reading instruction by testing these limits, which must be made explicit at the therapist's judgment. As the limits are accepted, they finally approximate those set up by an ordinary teacher-pupil situation.

In conclusion, reading therapy uses limits in much the same way as other therapies. It differs from them in having specific as well as general aims for its clients, in using reading as content about which to structure its processes, and in making use of a special milieu to facilitate these processes.

Received October 10, 1949.

REFERENCES

1. ALLEN, F. H. *Psychotherapy with children.* New York: Norton, 1942.
2. AXLINE, VIRGINIA M. *Play therapy.* Boston: Houghton Mifflin, 1947.
3. BIXLER, R. H. Limits are therapy. *J. consult. Psychol.*, 1949, **13**, 1-11.
4. BOND, G. L. AND HANDLAN, BERTHA. *Adapting instruction in reading to individual differences.* Minneapolis, Minn.: Univ. Minnesota Press, 1948.
5. DOLCH, E. W. *Group word teaching game.* Champaign, Ill.: Garrard Press.
6. ELLIS, A. Results of a mental hygiene approach to reading disability problems. *J. consult. Psychol.*, 1949, **13**, 56-61.
7. SMITH, M. An approach to the study of the social act. *Psychol. Rev.*, 1942, **49**, 422-440.
8. TAFT, JESSIE. *The dynamics of therapy in a controlled relationship.* New York: Macmillan, 1933.

A CLIENT-CENTERED APPROACH TO VOCATIONAL COUNSELING

G. D. BARAHAL, L. M. BRAMMER, AND E. L. SHOSTROM
STANFORD UNIVERSITY

THIS study grew out of the necessity to convert the services of the Stanford Veterans Guidance Center to meet the needs of the general student population. An earlier study at the Stanford Center [3] suggested that not all the Veterans Administration techniques were appropriate in the counseling of college students. Client-centered techniques and materials were developed which were felt to be philosophically and psychologically more compatible with the needs of students. This adapted counseling process was evaluated in a controlled experiment which compared two counseling processes — a traditional, highly-structured, counselor-centered approach and a client-centered plan. The latter plan included the adapted techniques and materials developed by the writers. The former approach included, among other things, a structured diagnostic first interview, counselor selection of tests, selection of specific vocational objectives, and considerable note-taking and form-filling on the part of the counselor. The latter plan included a precounseling group orientation, a permissive nondiagnostic first interview, client participation in test selection, new test interpretation materials, and other techniques described later.

PROCEDURE

The subjects were one hundred male, non-veteran, college freshmen volunteers divided equally into control and experimental groups. These groups were comparable in regard to age, scholastic aptitude, socioeconomic status, and amount of previous counseling. As the subjects appeared for counseling the fifty odds formed the control group; the fifty evens comprised the experimental group. They were assigned to three counselors who had received prior training and psychodramatic role-taking practice in the use of the two counseling approaches.

Veterans Administration forms and techniques were utilized with the control group, while the adapted client-centered materials and procedures were used with the experimental group. The complete counseling process required a total of three hundred hours of counseling time and five hundred hours of testing—equally divided between the two groups.

The groups were processed separately over a nine-week period in the following sequence:

1. First four weeks—the control group.
2. Fifth week—the counselors reviewed and practiced the adapted techniques and materials.
3. The last four weeks—the experimental group.

The two groups were processed separately because of the difficulty of counselor role-changing during the day, and the possibility that the subjects would discuss the differential counseling methods and materials with each other. It was felt that such discussion might invalidate the feeling-tone criterion which is described later.

THE ADAPTED TECHNIQUES USED WITH THE EXPERIMENTAL GROUP

The principal adapted techniques and materials[1] included a group orientation which was designed to give the client a realistic level of expectation prior to the first interview—an appreciation of what the client could expect from the counseling process, the strengths and limitations of tests, the role of the counselor and the client's role. In addition, at this group meeting a "General Information Form" was handed to the client for completion prior to the first interview with the counselor.

[1]The adapted materials described briefly here are being published by the Stanford University Press. All inquiries should be addressed to Stanford University Press, Stanford, California.

Having the "General Information Form" completed prior to the client's initial contact with the counselor, it was felt, would result in a permissive counseling interview rather than a probing, form-filling session. The use of this form, furthermore, might initiate an analysis and crystallization of the problem by the client *himself* before the first interview.

At the end of the first interview, the counselor and client *together* selected those tests which they felt could contribute something to the solution of the client's problem. The use of a "Test Selection Form," which included tests suitable at the college level, permitted this active participation by the client.

In the second interview, test results were interpreted to the client with new test interpretation forms which explained percentile scores, norm groups, and test prediction in meaningful terms.

Also, in this second interview, another new aid, a "Formulating Educational-Vocational Plans Summary," was used to encourage flexible vocational planning by the client instead of the selection of a rigid vocational objective as was done with the control group. During the counseling process the client was introduced to an additional self-directive learning device, a "Career Study Outline," which was designed to assist him in the gathering of significant occupational information.

The latter part of the final interview included the preparation of a "Career Planning Brochure" which, among other things, included the test interpretations, the "Career Study Outline," and the "Formulating Educational-Vocational Plans Summary." The client took this brochure with him as an aid in his future planning.

CONTROL TECHNIQUES AND MATERIALS

The control group clients were processed with the procedures and materials commonly used in Veterans Administration Guidance Centers and described in detail by Scott [10].

EVALUATION PROCEDURES

In order to obtain valid reactions to the counseling process, an independent interviewer conducted recorded interviews with all the clients in both groups on the evening following the completion of counseling. This interview was structured to the extent that it covered certain items; but considerable latitude was given to the interviewer to expand these questions. The interviewer covered twenty items which included the following: "As a result of this counseling program do you have a clearer conception of your future vocational plans?" "Do you feel that the tests helped you?" "Do you feel that the conferences with your counselor helped?" "Do you feel that the information in the Vocational Library of the Guidance Center helped you?" "Did the Center seem to operate efficiently and smoothly?" "Do you have any comments to make on the following people: the receptionist, the psychometrist, the counselor?" "How do you feel now about planning your *own* vocational future?" It was stressed that the client's identity would remain anonymous and that he could say anything he wished about the counseling service.

After these evaluation interviews were completed, three raters listened to all of the recordings and independently rated the twenty items. They used the following five-point scale:

1. Very Enthusiastic
2. Strong Satisfaction, but no enthusiasm
3. Satisfied, but with some reservations
4. Apathetic and/or mildly critical
5. Dissatisfied and/or highly critical

In addition to rating responses to specific questions, the evaluation interviewer and the three raters independently assigned an overall feeling-tone rating. They used the same five-point scale.

Judgments were based upon affective components of client responses. This criterion of client feeling tone, it was felt, was compatible with the client-centered approach in counseling. Other criteria such as stability of vocational choice and improvement in grades have been used in other studies [1, 2, 3, 4, 5, 6, 7, 8, 9, 11]. Although these are desirable outcomes of the counseling process, it was felt that evaluating the effectiveness of counseling in terms of the client's frame of reference had been neglected in previous studies. Does the client manifest feelings of satisfaction with the counseling process? Does he feel that he can resolve the problems which may arise in the future? In other words, the basic considerations in this study were not whether the client continued with his vocational objective, or changed it after counseling, or whether his grades improved; more important was the way the client *felt*

about remaining with or changing his objective and planning his own future.

The raters held several psychodramatic sessions to familiarize themselves with this feeling-tone criterion. They practiced distinguishing degrees of feeling tone to facilitate agreement upon the quality and intensity of the affective responses being rated. Correlations of the ratings revealed high reliability among the three raters. With the Pearson method, the coefficients ranged from .83 to .89. The correlation between the evaluation interviewer's overall rating and the average of the three independent raters' overall rating was .82.

Since information, as well as positive attitudes are important in future planning, a short achievement test was administered to both groups to determine what they had learned about their test scores, occupational information, and additional counseling facilities on the campus. This test as administered at the completion of the evaluation interview so that it would not influence the attitudes expressed during that interview. In addition, it was given to 127 Stanford freshmen who had not availed themselves of the Guidance Center Services.

RESULTS AND DISCUSSION

The critical ratio technique was utilized to test the significance of the differences between the ratings and test scores for the two groups. Major findings in regard to differences in feeling tone were as follows:

1. The difference between the mean overall ratings of feeling tone was significant beyond the one per cent level of confidence. The experimental group tended to be more satisfied with the counseling process than the control group.

2. The differences between the feeling tone ratings regarding the Testing Service, the Vocational Library, and the Counseling Service were significant at the one per cent level of confidence. The experimental clients felt more satisfaction with these services than the control clients. A typical verbatim comment concerning the tests, as expressed by an experimental group client, follows:

> They brought me out of a fog. They gave me a clearer understanding of where I was going. They showed that there are many other possibilities.

Another client commented:

> I didn't expect them [tests] to solve anything. They stimulated my thinking.

A typical control group reaction to the conferences with the counselor follows:

> Q: Do you feel that the conferences with your counselor helped?
>
> A: Considerably. The second interview was very helpful but the first one I was a little disappointed in. The few questions he asked (pause) I guess they had some bearing (pause) like going way back to grade school. (pause) I didn't find any necessity for that interrogation.

3. The experimental clients felt that they learned more about their interests and aptitudes, and that they were more sure of their future plans than the control clients. The differences between the feeling tone ratings on these questions were significant also at the one per cent level of confidence. These findings were interpreted as indications of greater self-direction on the part of experimental clients. A typical experimental client comment concerning his ability to plan his own future follows:

> Q: How do you feel now about planning your own educational and vocational future?
>
> A: I feel secure, I feel as if I'm headed on the right track—like I have something to walk on.

Another replied:

> I think I'm the only one who can do it really. It's my life. If someone told me I had to do it, I wouldn't have done it. It [counseling] has sure helped.

4. Differences between the groups concerning how they felt about the administration of the Center, and the clients' overall estimates concerning the monetary value of the service revealed no significant differences between the groups. Both groups revealed equal enthusiasm and strong satisfaction with the operational efficiency of the Guidance Center.

The mean dollar estimate of the value of the guidance service, as expressed by the control group, was $37.13; the experimental group, $46.20. The actual cost for tests and personnel time per client was approximately $21. However, the experimental group expressed more positive feeling tone in regard to the worth of the service. This feeling-tone difference was

significant at the one per cent level of confidence.

5. Differences between the mean ratings of Guidance Center Personnel—the receptionist, the psychometrist, the counselors — were significant beyond the one per cent level of confidence. The more permissive and accepting techniques utilized with the experimental group resulted in more positive feeling tone toward the participating Guidance Center personnel than the more rigid procedures used with the control group.

Client perceptions of the counselor appeared to be influenced by the techniques he used. Control clients exhibited ambivalent feelings. It is quite probable that with the control clients the counselor's attitudes were sufficiently permissive, but that his techniques were not in harmony with this permissiveness. With the experimental group, when counselors were permissive both in their attitudes and the techniques they employed, there was a more positive feeling tone toward them as individuals. An example of an experimental client comment follows:

Q: Do you feel that the conferences with your counselor helped?
A: I enjoyed them very much.
Q: Would you care to comment further on this?
A: I liked them because he [the counselor] guided my thinking rather than telling me what he thought.

The results of the Vocational Guidance Test, which was administered to the clients in both groups as an additional criterion, follow: 1. The difference between the mean scores of the control and experimental groups was significant at the four per cent level of confidence. 2. The difference between the mean scores of the experimental group and the mean scores of the noncounseled freshman group was significant beyond the one per cent level of confidence. The information criterion did not seem to be useful in this instance for differentiating between the two counseled groups. However, the highly significant difference in scores between the counseled and noncounseled groups illustrates the learning which takes place in the counseling process.

CONCLUSIONS

The following conclusions appear warranted:

1. Client-centered permissive counseling procedures and materials, as employed with the sample in this study, resulted in greater client satisfaction with the counseling process than highly-structured counselor-centered procedures.

2. Information techniques (tests, occupational information, and personal histories) can be used effectively within a permissive client-centered approach to counseling.

3. Client satisfaction (manifested through feeling tone and evaluation interview comments) appears to be a valid criterion for evaluating counseling, since the client apparently considers counseling successful if it reduces his anxiety and satisfies the needs which he brings to the counseling situation.

4. Client-centered techniques and materials seemed to result in greater client feelings of self-direction than the more rigid control processes and materials.

This study has provided Stanford University with an effective plan for adapting the services of its Veterans Guidance Center to meet the needs of its general student population. Perhaps other guidance centers will find the techniques and materials useful in their future counseling programs. It is hoped, further, that this study will stimulate additional research so vital to the improvement of guidance practice.

SUMMARY

This study, initiated in the spring of 1949, undertook to investigate the desirability of extending the services of the Stanford Veterans Guidance Center to the general student population, develop client-centered and informational techniques which would meet the needs of Stanford students, veterans and non-veterans alike, and evaluate the effectiveness of these client-centered techniques and materials.

Control and experimental groups, composed of one hundred volunteer Stanford freshmen, were used to compare the effectiveness of these newly developed client-centered techniques and materials with the structured materials and procedures commonly used in Veterans Administration guidance centers.

A total counseling and testing program was made available to both groups. VA procedures were utilized with the control group; the new client-centered materials and procedures with

the experimental group. At the completion of the counseling process, an independent interviewer conducted a twenty-minute recorded interview with all one hundred clients. Three independent raters listened to the recordings and rated each of the twenty evaluation interview items on a five-point feeling-tone scale. Feeling tone (client affective expression of satisfaction with parts of and the total counseling process) was the criterion used to evaluate the effectiveness of counseling.

Results indicated that the client-centered techniques and materials resulted in more positive client satisfaction with counseling, and more self-direction.

Received March 31, 1950.
Early publication.

References

1. Brown, M. T. The veteran reports one year later. *Occupations*, 1947, **25**, 209-212.
2. Brown, M. T. The veteran reports two years later. *Occupations*, 1948, **26**, 364-366.
3. Corcoran, Mary E. Factors affecting the stability of the vocational plans of Stanford student veterans. Unpublished Master's thesis, Stanford University, 1948.
4. Deck, A. O. and Reeves, P. Effects of advisement upon continuation of training under Public Law 346. *Sch. & Soc.*, 1948, **67**, 429-431.
5. Froehlich, C. P. *Evaluating guidance procedures.* Publication, Miscellaneous 3310, Washington: United States Government Printing Office, 1949.
6. Kircheimer, Barbara A., Axelrod, D. W. and Hickerson, G. X., Jr. An objective evaluation of counseling. *J. appl. Psychol.*, 1949, **33**, 249-257.
7. Lipsett, L. and Smith, L. F. The Rochester Veterans Guidance Center takes stock. *Amer. Psychologist*, 1948, **3**, 12-15.
8. Long, L. and Hill, J. A follow-up study of veterans receiving vocational advisement. *J. consult. Psychol.*, 1947, **11**, 88-92.
9. Newhouse, G. L. The effectiveness of the San Francisco Veterans Counseling Center. Unpublished Master's thesis. Stanford Univ., 1948.
10. Scott, I. D. *Manual of advisement and guidance.* Washington: United States Government Printing Office, 1945.
11. Ward, J. R. and Tyler, Leona E. A preliminary report of an evaluation of the Veterans Administration counseling service in the University of Oregon. *Amer. Psychologist*, 1947, **2**, 416.

A FACTOR ANALYSIS OF RORSCHACH SCORING CATEGORIES

J. R. WITTENBORN
YALE UNIVERSITY

THE present study was undertaken for the purpose of testing three hypotheses. These hypotheses were developed during the course of a series of investigations concerned with the use of the Rorschach test.

Several of the analyses which have been published in the present series of studies [4, 5] provide evidence that the traditional distinction between Rorschach responses based primarily on the perception of human movement, and Rorschach perceptual responses based on an employment of color, is congruent with the manner in which various manifestations of these two classes of responses are intercorrelated. In this respect, at least, the universally observed scoring distinction between the movement and color responses is justifiable. The distinctions between the movement and color responses provided by the various interrelationships examined were relative and far from absolute, however. Inasmuch as there is a variety of different Rorschach determinant scores, it is possible that the distinctions revealed between the human movement and color scores are minor relative to the distinctions that exist between some of the other determinant scores. This, however, is not implied in the Rorschach literature. As a matter of fact, in most writings the distinction between human movement and color receives a primary emphasis. Accordingly, one would not predict that this distinction would be obscured by other marked and broadly relevant distinctions between Rorschach determinants. The magnitude of the distinction between the human movement response score and the color response score may be evaluated by comparing it with the magnitude of the distinctions which exist between other classes of response. Such a comparison is relevant to Rorschach theory and may have some practical implications.

Under certain conditions, a factor analysis of the intercorrelations among various commonly employed Rorschach scoring categories should reveal the difference between the human movement and the color responses. This possibility is stated formally in Hypothesis I.

Hypothesis I: In a factor analysis of the various Rorschach response categories, the pattern of factorial composition for the human movement response will be different from the pattern of factorial composition for responses involving the use of color.

In order for the factor analysis to be relevant to Hypothesis I, it is necessary for both the human movement response score and some of the color response scores to show large correlations with several scoring categories, i.e., it is necessary for both human movement and color scores to have an important amount of common factor variance. If, under this condition, the analysis does not reveal evidence for Hypothesis I, the distinctions between the human movement and the color responses which have been revealed in earlier analyses may be considered to be minor relative to the distinctions which exist between other response categories.

As a result of a few different fragmentary clues provided by earlier analyses [3, 5], it was inferred that some of the Rorschach responses differed from each other with respect to the degree of perceptual control characterizing them. Although the exact nature of this "control" was not and is not clearly specified, human movement responses and large detail responses were conceived as involving a high order of perceptual control, whereas whole responses and color-form responses were considered to be alike[1] in that they involve a flex-

[1]To what degree the suspected behavioral similarity between color-form and whole responses is due to a possible prevalence of whole color-form

ible perceptual approach requiring appreciably less perceptual discipline. A low order of perceptual control is tentatively conceived to be manifested by responses which are incautious, possibly spontaneous or impulsive, to a degree which results in a relative disregard for or an unawareness of the purely formal, literal, or concrete response possibilities. In order to apply this conceptualization broadly, the pure color, the pure texture, and the pure diffusion response categories may be added to the whole and the color-form response categories. Similarly, in order to broaden the tentatively conceived class of responses which expresses a high order of perceptual control, the pure form responses and the other detail response categories are added to the human movement and the large detail response categories. This conceptualization requires that the pattern of intercorrelation among the response categories should be organized in the manner specified by Hypothesis II.

Hypothesis II: The whole, color-form, pure color, pure texture, and pure diffusion response categories are intercorrelated in a manner which yields a factor different from the factor which is most important in determining the common factor variance of the pure form, human movement, and the various detail response categories.

A classification of responses on the basis of the degree to which they require perceptual control is closely akin to the common practice of evaluating responses on the basis of the degree to which they reflect spatially definite well-conceived formal percepts. The conceptualization responsible for Hypothesis II is somewhat different from the usual form-level evaluation in that it is intended to emphasize distinctions in perceptual approach rather than distinctions in perceptual quality or plausibility. It also makes no provision for the qualitative distinctions among the texture, diffusion, and color determinants. It implies that the distinctions between responses which are based upon the degree of perceptual control transcend the distinctions among the different classes of determinants. Hypothesis II, if a true hypothesis, may invite a shift in emphasis from the qualitative distinctions between certain classes of determinants to a system of distinctions based on perceptual control; it does not preclude, however, that there are discernible systematic distinctions among whole, uncontrolled color, uncontrolled texture, and uncontrolled diffusion responses.

Both as a result of clinical applications of the Rorschach method and researches concerning the psychological significance of Rorschach responses, the writer has come to suspect that although the general productivity of the patient greatly influences the score for most response categories, it does not influence the score for the various response categories equally. If productivity is an important determiner of the frequency with which responses fall in the various response categories, the possibility exists that most of the reliable variance of some of the scoring categories is due primarily to the general productivity of the patients. For the purposes of the present experiment, this possibility is formally stated by Hypothesis III.

Hypothesis III: The Rorschach scoring categories will be correlated with the total number of responses (R) in such a manner as to result in a general or quasi-general factor of productivity.

If Hypothesis III is supported by the present data, it becomes possible to examine the validity of Hypothesis IV which has practical implication for the use of the Rorschach test.

Hypothesis IV: Response scoring categories differ greatly in the degree to which their common factor variance is due to a productivity factor.

Hypothesis IV is worth examining because the degree to which the various scoring categories are independent of general productivity is unknown, and if the various scoring categories differ markedly in this respect, it would suggest, all other things being equal, that the scoring categories most independent of productivity be accorded a more uniquely significant meaning than those which reflect primarily the productivity of the patient, i.e., the recognition of response categories which are primarily measures of productivity would contribute to efficient and economical scoring procedures and reduce unrecognized sources of error necessar-

---

responses to one or two cards is unknown, but the concept of uncontrolled perceptual responses as expressed in Hypothesis II is extended beyond such specific possibilities and the concept must stand or fall on its broad implications.

TABLE 1
DEFINITIONS OF SCORING CATEGORIES

Variable No.	Symbol	Definition
1	W	Whole blot
2	D	Large usual detail
3	d	Small usual detail
4	Dd	Unusual detail
5	S	White space
6	M	Figures in human-like action (human, mythological, or animal)
7	FM	Animals in animal-like action
8	m	Abstract or inanimate movement
9	K	Shading as diffusion (smoke, clouds)
10	FK	Shading as three-dimensional expanse in vista or perspective
11	F	Form only, not enlivened
12	Fc	Shading as surface appearance or texture, differentiated
13	c	Shading as texture, undifferentiated
14	C′	Achromatic surface color
15	FC	Definite form with bright color
16	CF	Bright color with indefinite form
17	C	Color only
18	P	Popular responses
19	O	Original responses, found not more than once in one hundred records
20	R	Total number of responses

ily inherent in practices which involve the multiple employment of the same variable which unrecognized appears in different guises.

SUBJECTS AND PROCEDURES

As a means of testing the three hypotheses, 21 different Rorschach scores were intercorrelated for a sample of 92 Yale undergraduates and the intercorrelations were submitted to a centroid analysis. The Rorschach scoring categories employed are shown in Table 1, and their intercorrelations in Table 2. The variables may be recognized as comprising the principal scoring categories of the Klopfer system. The intercorrelations were computed by the Pearson product-moment method, which requires that the variables be linearly related to each other. An examination of numerous scatter diagrams between a variety of scoring categories gave no indication that the relationships were nonlinear. All of the distributions of the scoring categories were positively skewed, however. Had the skewness in some instances been negative, it is possible that the linearity of some of the interrelationships among the scoring categories would have been distorted. Some of the scores occurred infrequently and were not found in every protocol. It is not customary to apply the product-moment coefficient to variables having such a small range[2]; nevertheless,

[2]In general it may be observed that the variables that have the smallest range tend to have relatively small correlations. This is not surprising, but this difficulty cannot be resolved in any simple manner. Since all the correlations are based upon the same sample, the usual considerations regarding heterogeneity of sample are not applicable nor is it clearly appropriate to regard the difficulty as a result of broad categories because an attempt to correct the variables in question for a broad category effect would involve ascribing statistically a refinement and sensitivity which actually does not characterize the scores. Since the present investigation is concerned with the conceptual implications of the Rorschach *scoring procedures*, it was

TABLE 2
INTERCORRELATIONS AMONG SCORING CATEGORIES

No.	Symbol	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	W																			
2	D	−194																		
3	d	−228	669																	
4	Dd	−138	608	749																
5	S	−143	464	512	660															
6	M	−041	498	560	665	616														
7	FM	−058	625	429	381	288	359													
8	M	115	403	351	297	287	379	375												
9	K	255	330	246	259	155	269	281	425											
10	FK	203	275	238	221	232	154	201	240	411										
11	F	−052	693	705	762	635	438	285	202	147	198									
12	Fc	100	587	583	605	519	579	326	450	349	153	460								
13	c	144	138	116	108	097	113	136	182	310	193	108	191							
14	C′	016	466	510	503	508	532	345	468	310	114	377	546	076						
15	FC	023	541	429	457	374	342	361	563	181	028	305	571	−084	464					
16	CF	410	241	090	114	−008	−108	−025	113	353	285	179	117	251	089	133				
17	C	170	178	072	137	066	−020	075	098	259	270	039	098	261	109	177	235			
18	P	−095	208	266	163	034	177	221	363	198	−005	025	357	187	206	356	−033	216		
19	O	−011	436	570	579	565	624	303	456	305	135	415	646	181	379	550	046	172	659	
20	R	093	879	783	813	598	629	581	478	429	379	797	710	195	551	584	315	224	188	551

TABLE 3
CENTROID FACTORS

No.	I	II	III	IV	h^2
1	087	472	134	—417	.422
2	788	—291	159	094	.740
3	745	—353	085	184	.721
4	773	—378	122	079	.762
5	629	—341	—049	—038	.516
6	656	—301	—191	094	.566
7	540	—127	—013	196	.346
8	601	151	—296	—122	.487
9	521	409	101	053	.452
10	383	277	323	096	.337
11	664	—385	350	—071	.717
12	765	—098	—166	—120	.637
13	284	375	144	197	.333
14	649	—122	—225	—137	.505
15	613	—102	—308	—289	.565
16	283	400	461	—274	.528
17	274	358	148	110	.237
18	384	261	—456	308	.518
19	744	032	—466	109	.784
20	941	—216	277	—064	1.013

TABLE 4
TRANSFORMATION MATRIX*

	I	II	III	IV
I′	556	104	—721	—403
II′	195	586	—197	761
III′	316	635	537	—425
IV′	744	—494	349	281

*In one of the rotations .709 was erroneously substituted for .707 in the sine and cosine. For this reason the rotated factors are not perfectly orthogonal (the departure from orthogonality is quite minute) and the communalities in Table 5 do not correspond perfectly with those of Table 3.

it is felt that a system of intercorrelations such as provided by Table 2 permits an evaluation of the hypotheses in question.

The four factors yielded by the centroid analysis are presented in Table 3. The factors in Table 3 may be rotated to the position described in Table 5 by an employment of the transformation matrix shown in Table 4. All of the various rotations were selected by an application of the usual criteria, i.e., maximizing the number of zero loadings and minimizing the negative loadings. The test clusters responsible for rotated factors II′, III′, and IV′ are relatively independent of each other. The clustering tendency responsible for factor I′ is not wholly distinct from that responsible for factor IV′. If the present study were concerned

decided to study the scores directly with whatever imperfections they may possess and not attempt statistically to refine them or their intercorrelations.

TABLE 5
ROTATED FACTOR MATRIX

No.	Symbol	I′	II′	III′	IV′	h^2
1	W	169	—015	576	—239	.420
2	D	255	023	110	812	.737
3	d	242	061	—021	810	.719
4	Dd	271	—035	036	827	.760
5	S	365	—007	—028	609	.505
6	M	433	061	—127	597	.564
7	FM	217	183	—001	515	.346
8	m	612	171	179	235	.491
9	K	238	361	456	236	.450
10	FK	—030	246	429	288	.328
11	F	105	—220	183	786	.711
12	Fc	583	033	141	526	.637
13	c	014	397	321	132	.278
14	C′	565	—005	065	426	.505
15	FC	669	—100	086	318	.566
16	CF	—024	—010	707	097	.510
17	C	039	318	347	109	.235
18	P	445	552	—089	084	.518
19	O	709	339	—041	406	.784
20	R	327	292	336	885	1.088

with factors per se, it would be necessary to rotate factors I′ and IV′ obliquely with respect to each other and it would be possible to make other minor improvements in the rotated factors.

From an examination of Table 5 it may be seen that, with certain qualifications, Hypothesis I is a true hypothesis. Variable 6, the human movement variable, draws most of its common factor variance from factor IV′; it also draws a substantial quantity of its factor variance from factor I′. The factorial composition of variable 6 is quite different from that of variables 16 and 17, the color-form and the pure color response variables; most of their common factor variance may be explained in terms of factor III′. On the basis of the present analysis, however, the form-color response variable (number 15) is more similar in its factorial composition to the human movement variable than it is to the other color variables. The factorial composition of the form-color response is not at all similar to the factorial composition of the color-form and the pure color responses. The form-color response is different from the human movement response in that it draws most of its common factor variance from factor I′ and is only secondarily dependent on factor IV′, which is the chief source of the common factor variance of the human movement response variable. It is interesting to ob-

serve that the factorial composition of the form-color response is quite similar to the factorial composition of the achromatic color responses (variable 14), the form-texture responses (variable 12), and to a lesser degree the inanimate movement (variable 8) and the human movement (variable 16) responses. This pattern of relationships is scarcely predictable from the Rorschach literature and the scoring and interpretive practices of Rorschach workers. The present sample suggests that the superficial logical similarity between the form-color and the color-form responses may have little behavioral basis. If the pattern of relationships revealed in the present sample can be verified in other samples, it may be desirable to consider a revision of our current beliefs concerning the behavioral or personality significance of the form-color response.

An examination of factor III′ shows it to be the factor required by Hypothesis II. Factor III′ may be seen to be determined by the classes of response (whole, pure vista, pure texture, color-form and pure color) which were judged by the writer on a priori grounds to be similar in that they presumably require a loose, informal, spontaneous perceptual approach not governed primarily by the most obvious spatial characteristics of the ink blot. Factor III′ is also seen to be correlated with variable 10, the form-vista variable; although no prediction concerning variable 10 was made. Variable 20, total productivity (*R*), participates to some degree in factor III′. This is not an untoward result and obviously does not detract from the meaning of factor III′.

A most impressive feature of Table 5 is its failure to correspond with the qualitative classifications of Rorschach determinants [movement (*M, FM, m*), vista (*K, FK*), texture (*c, Fc*) and color (*C, CF, FC*)] which are emphasized so insistently in the literature and about which elaborate and tediously arbitrary scoring distinctions have been devised. For example, the factorial compositions of the form-vista, pure vista, pure texture, and pure color response variables are strikingly similar to each other and involve primarily factors II′ and III′. The form-color and form-texture response variables have a highly similar factorial composition which does not involve factors II′ and III′, but instead draws upon factors I′ and IV′. Thus, we see that qualitatively similar Rorschach response determinant variables not only fail to cluster together in a manner to form factors, but they have dissimilar factorial compositions and responses which fall in one particular qualitative determinant classification are inclined to resemble most certain responses which belong to some other class of determinants. Findings such as these suggest that the behavioral or personality significance traditionally ascribed to many of the specific Rorschach scoring categories may be incorrect in its emphasis even if valid to a degree.

Factor IV′ seems to satisfy the requirements of Hypothesis III. This would be more obvious if factors I′ and IV′ were rotated obliquely with each other, or if the reference vector corresponding to factor IV′ were projected through variable 20. Nevertheless, Hypothesis IV may be discussed in terms of factor IV′. The relative importance of general productivity may also be discerned by an examination of the intercorrelations, Table 2. For the most part, the variables which are important to factor IV′ also have substantial loadings with factor I′ and in general these variables have a highly similar factorial composition; specifically all of the various detail scores are highly similar to each other and to a striking degree are similar in factorial composition to all of the scores which are based on responses which primarily emphasize form. At any rate, either the correlations with variable 20 (total number of responses, given in Table 2) or the loadings of the various scores on factor IV in which variable 20 is so important may be taken as evidence that many of the Rorschach response scores are for the most part highly similar to each other and are also a measure of the productivity of the subject in the Rorschach situation. All of the various types of scores are not equally affected by the general productivity of the patient. Some of the Rorschach scores, particularly those which have high loadings with factors II′ and III′, show important evidence of having a behavioral significance quite different from the behavioral significance of the general productivity of the subject.

It may be noted that variable 18, the number of popular responses, is less important in the productivity factor than variable 19, the number of original responses. This probably

arises from the fact that there is a limited number of popular responses; because of their popularity the percentage of popular responses for moderately productive subjects would be as great or greater than the percentage of popular responses for highly productive subjects. Similarly, since originality of responses is based on a frequency of occurrence criterion, the subjects producing the largest number of responses, all other things equal, have a greater chance of producing original responses.

In general, the manner in which variables are intercorrelated depends upon patterns of differences among the individuals comprising the sample. The pattern of differences among the individuals comprising the present sample is in some respects reminiscent of clinically familiar personality patterns. This is particularly true in the case of the variables which cluster together in a manner to form factor III′. The tendency to give numerous whole responses and to give responses in general which are spontaneous, impressionistic, and not exclusively formal and literal in their spatial aspects has often been found to characterize the so-called hysterical personality. Such perceptual tendencies on the Rorschach are also often considered to be characteristic of the classical personality trait of extroversion. To what degree either the hysterical personality or the extrovert personality produces a relatively large number of the kinds of Rorschach responses which determine factor III′ would not be difficult to determine. As a matter of fact, it would be very simple from the arithmetic standpoint to weight the types of responses contributing to factor III′ in a manner to yield a scale of the response tendency or characteristic responsible for the factor. If this were done, the behavioral significance of this scale could be explored by correlating it with results of other personality tests and with different clinical and behavioral criteria.

The Rorschach responses responsible for factor IV′, viewed as a group, are also reminiscent of a familiar personality reaction pattern which has considerable clinical interest. Specifically, Rorschach protocols produced by ruminative, obsessive, compulsive individuals are commonly characterized by a preponderance of detailed responses, relatively numerous human movement, animal movement and pure form responses, a high level of productivity and a tendency to avoid informal responses such as those contributing to factor III′.

The participation of *W* responses in factor III′ may be puzzling to some readers and they may be inclined to suspect that its presence in factor III′ is due to a tendency for *K, c, CF,* and *C* to be determinants for whole responses. To what degree this is true cannot be established from the present analysis but in an earlier study [3] several whole responses to achromatic cards were found to be highly correlated with the *CF* responses to chromatic cards 8 and 9.

CONCLUSIONS

As a means of evaluating certain hypotheses concerning the behavioral similarities and dissimilarities of Rorschach responses, 21 of the basic scores of the Klopfer system were intercorrelated and submitted to a centroid analysis. The intercorrelations could be accounted for in four factors and after a series of orthogonal rotations several clustering tendencies became conspicious. As a result of this analysis and within the limitations of the sample of 92 students employed, the following conclusions are offered:

1. The factorial composition of the human movement response category is distinctly different from the factorial composition of the color-form and the pure color response categories. The color-form and the pure color responses have a highly similar factorial composition.

This result combined with the results of earlier studies [4, 5] comprises considerable justification for the traditional distinction (between human movement and the color responses) which is emphasized in the scoring of the Rorschach test and in an interpretation of Rorschach protocols.

2. The factorial composition of the form-color response category is quite different from the factorial composition of the color-form and color response categories. This result suggests that the common practice of regarding the three classes of color responses as similar in their implications for the affective features of human behavior should be viewed with reservations. In the present study, the factorial composition of the form-color response cate-

gory is more similar to the factorial composition of the human movement response category than it is to the other color response categories.

3. Factor III′ in the present study fulfills the requirement of Hypothesis II which states, "The whole, color-form, pure color, pure texture, and pure diffusion response categories are intercorrelated in a manner which yields a factor different from the factor which is most important in determining the common factor variance of the pure form, human movement, and the various detail response categories." The evidence for the validity of Hypothesis II suggests that it may be profitable to employ a new Rorschach scoring principle which is based on a weighted combination of the response categories which contribute to factor III′

4. Factor IV′ may be offered in support for both Hypothesis III ("The Rorschach scoring categories will be correlated with the total number of responses [*R*] in such a manner as to result in a general or quasi general factor of productivity") and Hypothesis IV ("Response scoring categories differ greatly in the degree to which their common factor variance is due to a productivity factor"). There are two possible practical implications for factor IV′:

(a) Factor IV′ may be produced by the obsessive, compulsive personalities in the sample and when its contributing response categories are appropriately weighted, it may provide a scale which has some clinical value.

(b) This factor has implications for the current use of the Rorschach test inasmuch as it implies that some of the separately interpreted scoring categories overlap each other and should receive relatively less weight than they now do.

5. In general, the present study suggests that an incorrect emphasis may have influenced the development of current Rorschach scoring procedures and interpretive practices. For example, many of the scoring categories which belong to the various broad classes of determinants, e.g., color, texture, or diffusion have a quite dissimilar factorial composition and in general the manner in which the various determinant scoring categories cluster together could not be predicted by an employment of the usual beliefs concerning behavioral implications of determinants.

Received September 9, 1949.

References

1. Klopfer, B., and Kelley, D. M. *The Rorschach technique.* Yonkers, N. Y.: World Book Co., 1942.
2. Wittenborn, J. R. Certain Rorschach response categories and mental abilities. *J. appl. Psychol.*, 1949, **33**, 330-338.
3. Wittenborn, J. R. A factor analysis of discrete responses to the Rorschach ink blots. *J. consult. Psychol.*, 1949, **13**, 335-340.
4. Wittenborn, J. R. Statistical tests of certain Rorschach assumptions: Analyses of discrete responses. *J. consult. Psychol.*, 1949, **13**, 257-267.
5. Wittenborn, J. R. Statistical tests of certain Rorschach assumptions. The internal consistency of scoring categories. *J. consult. Psychol.*, 1950, **14**, 1-19.

PSYCHOLOGICAL CONSTRICTION ON SEVERAL PROJECTIVE TESTS [1]

FRANCES MERCHANT CARP
WESTERN MICHIGAN COLLEGE OF EDUCATION

PROJECTIVE methods have gained widespread use in recent years as clinical application has shown their diagnostic and prognostic usefulness, but this reputation from clinical practice continues to be their main source of evidence, while corroboration by systematic attention to validity and reliability remains inadequate. The continuing dearth of objective evidence concerning projective techniques may be due partly to the inaccessibility of a basic assumption. When projective materials are used in evaluating personality it is hypothesized that the overt behavioral data (test responses) represent covert personality factors. Through this assumption, access is claimed to new areas or levels of personality, but proof is difficult since one member of the postulated relationship is not open to observation.

Since it seems impossible to make direct tests of relationships between projective materials and personality variables, it is desirable to obtain the approximations by indirect methods. The comparison of projective test results with other behavior is one way to get information concerning their validity. Swift [10] used this method when she compared insecurity as evaluated by the Rorschach test to insecurity as indicated by other behavior clinically judged to be symptomatic.

A second indirect way to investigate projective test validity is to compare the results of different tests. This method has been used in the development of intelligence tests [9] and it was used by Magaret and Simpson [8] in a recent investigation of measures of deterioration in psychotics.

[1]Grateful acknowledgment is made to Dr. M. M. James, Dr. E. R. Hilgard and Dr. C. P. Stone, all of Stanford University, for their helpful advice.

Neither of these methods identifies the fallacious member, in case of disagreement between test and another criterion or between two tests. However, in the absence of a direct approach it seems worthwhile to obtain what information these indirect ones can furnish.

Though clinical psychologists using projective tests have little difficulty in coordinating and harmonizing results and interpretations from several tests, this is not satisfactory evidence of intertest agreement since clinicians commonly view together all test results and other information regarding the individual. Undeniably there is unique value in the clinician's synthesis of all available material, in what Frank calls "the method of biological relativity" [3, p. 4]. However, that practice does not decrease the importance of developing to its finest possibility every tool available to the clinician for the formulation of his diagnosis. An initial step in such development is the realistic evaluation of the tools now used. An analysis of the unique contribution of each tool seems essential to such evaluation. To provide evidence of intertest agreement, each must be interpreted independently and then the order of agreement ascertained.

Similarly, agreement as to the interpretation of a particular projective test is of questionable value when that agreement is produced by a conference method. More data are needed on the relationships when the same test is interpreted independently by several psychologists. Since the interpretation of projective materials is criticized as esoteric and individualistic and therefore not amenable to scientific development, it is important to obtain information concerning the reliability with which the tests can be evaluated. Reliability of interpretation is, of course, prerequisite to test reliability and validity, and it is suit-

able for concurrent study. Unlike the latter, it can be attacked directly through comparison of independent interpretations of qualified workers.

THE STUDY

Evidence concerning the validity of some projective tests now in use as clinical tools was sought through two indirect methods: (a) by testing the agreement among various projective tests and (b) by testing the agreement of the projective tests with certain other behavior. To obtain information concerning the reliability of the scores upon which the tests for validity were based, the reliability of interpretation of the projective materials was also studied.

Because children were used as subjects, the projective methods selected as suitable were the Rorschach test, crayon drawing, and play construction. Constriction was the psychological variable studied.

Specifically, then, the study sought information concerning: (a) the agreement among the Rorschach test, crayon drawing and free-play construction on psychological constriction, (b) the agreement of each of the projective tests with certain other criteria of constriction and (c) the agreement between clinical psychologists' independent evaluations of the same projective materials.

Subjects. The entire third grade of a public school served as subjects. There were 47 boys and 49 girls. Educational Quotients as measured by the Progressive Achievement Test, Primary Battery, Form C, were available for 88 of the 96 children. A distribution of these scores is given in Table 1.

TABLE 1
EDUCATIONAL QUOTIENTS

E.Q.	f
120	2
115	3
110	2
105	12
100	14
95	15
90	11
85	14
80	8
75	4
70	1
65	1

Procedure. The first contact made with the children was in their classrooms. The experimenter, a stranger to all of them, was introduced to the children by the classroom teacher. Each child was given a full set of crayons, a pencil and one sheet of 12 by 18 inch drawing paper, and was directed to draw his "own family." As soon as these papers had been collected, others were distributed and new instructions were given to draw "your very best wish."

The writer administered the Rorschach test individually to each subject in a quiet and secluded room. No sample blot was used and the order of procedure was that used for adults, with the performance proper completed for all ten cards before the inquiry was begun. This seemed to pose no problem for the children and none was reluctant to go back through the cards for the inquiry.

Individual play sessions were conducted by the writer. On a low table was a construction board 40 inches by 24 inches in size, with a backboard 20 inches high.

Sitting just inside the door and at the child's back and right as he stood at the table, the experimenter gestured to the play materials displayed in full view and said, "Here are some things people like to play with. See, here are blocks if you want to build, and here is furniture, and airplanes and boats and cars, and animals, and here are the people. You can use any of the things you want to, and do anything with them. While we are here, this room is yours and everything in it is yours. You can do whatever you want to. For a while I will go ahead with some work I have to do, [picking up notebook and pencil] and you go ahead with whatever you want to do." The experimenter began writing in the notebook.

The play session ended whenever the child seemed to come to a stopping place. Usually this was clearly defined by his remarking, "It's all fixed," or exclaiming, "There!" or by asking the experimenter to look at what he had done. A few children simply stopped playing and stood quietly until the experimenter took notice of them. Some began putting the toys away. The experimenter interrupted four of the girls when it became obvious that their housekeeping play had settled into a routine to be repeated over and over.

The final contact was in the classrooms. Again crayons, pencil and paper were supplied, and this time the children were instructed to

draw "how you feel." All testing was completed within one month.

Nonprojective Data. Their classroom teachers rated the children on the Trait Rating Blank devised by Maud A. Merrill using traits defined by Jean MacFarlane. The Blank is made up of 19 items, each to be marked on a linear scale with five behaviorally defined points. To avoid biasing the results, the teachers were asked to fill out the entire blank rather than selected items, though only three of the traits were used in the study. Those considered possibly concurrent with psychological constriction were Shy-Bold, Deliberate-Impulsive and Unsocial-Social. The defined points on the scales are:

Shy-Bold: "(1) Extremely retiring. Avoids all possible social situations. Forcing situation causes acute discomfort. (2) Retiring but not so extreme as 1. Avoids social situations often, but not on every possible occasion. Slight discomfort if general attention is focussed on child. (3) Neither shy nor bold — average — does not retire from, nor seek social situations. (4) Likes attention but not so extreme as 5. Is at east in social situation. (5) Plunges into any social situation. Glories in general attention. Does things designed to attract attention."

Deliberate-Impulsive: "(1) Can't be budged — does nothing without cousidering it from every angle, carries this type of reaction to the extreme of small unimportant details. (2) More deliberate than average, does very little on the spur of the moment. Rather inclined to ponder on possible results of a course of behavior. (3) About average, may do things on the spur of the moment but on the whole foresees consequences of a certain mode of behavior, and acts accordingly. (4) Impulsive behavior but not so extreme as 5. Frequently gets into situations because of impulsive action. (5) Erratic behavior — never considers consequences of act performed — always rushing headlong into situations — completely thoughtless."

Unsocial-Social: "(1) Very unsocial. Works, plays, studies alone. Repulses advances of other children either from excessive shyness or preference. Refuses to enter into any kind of social activity of children. Rude. (2) Prefers solitary play but can adjust to social situations on occasion. (3) Likes both social and solitary situations. No particular preference. Not a leader nor outstanding. Has own friends. (4) Prefers social play. Very friendly and outgoing, has lots of friends. (5) Likes practically everybody. Is a leader in social activities. Extremely outgoing. Wants to be in a crowd every minute."

The teachers reported confidence in their ability to mark this scale, which deals in concrete behavioral characteristics meaningful to the psychological layman and open to a teacher's observation.

A sociometric method was used to obtain another indication of seclusiveness, since it is a very probable behavioral correlate of psychological constriction. An almost invariant and early symptom is paucity of social contacts and lack of interest in other people, especially contemporaries. In an individual interview, each child in the study was asked to name the children he would like to sit by in school. After each response he was asked, "Anyone else?" until he said that was all. The number of children he named was used as an index of each child's sociability.

Treatment of the Data. The test results were coded and all other identifying material was removed. In order to change the drawings as little as possible by the coding, butcher's tape was pasted over any identifying material and the nature of that material was written or printed on the tape, in the color used by the child. For instance, a red printed "Name" was pasted over a red printed "Carolyn."

All persons who worked on the data were staff members in clinical psychology or advanced graduate students in clinical psychology at Stanford University.[2] All had suitable training and experience with the techniques.

More Objective Scores. For each technique some fairly objective scores were obtained as possibly indicative of constriction. *F%* was used for the Rorschach test. Split-half reliability (odd-even numbered cards) of *F%* was .86.[3] The consistency of the two scores is indicated by 95 per cent agreement on the determinants of the 13,670 main responses.

Scores for the drawings were (a) the number of colors used and (b) the area of the page used. The first, obtained by direct count, involved little scoring variability. For the second, correlations between rankings by two psychologists were .94 for the "family," .94 for the "wish" and .99 for the "feel" drawings.[4]

[2]Abraham Carp, Wm. A. Cass, Jr., Erik Erickson, Lucy Lawson, Maud A. Merrill and Frances Orr were the scorers and raters who made the study possible.

[3]Product moment correlation corrected by the Spearman-Brown formula. The uncorrected value was .76.

[4]Rank-order correlations.

Although both indices were relatively stable on a score-rescore basis, neither was consistent from drawing to drawing. Interdrawing correlations for the number of colors .20, .19 and .15,[5] and those for area of the page were .29, .04 and .08.[6] Because of this lack of consistency each set of drawings was treated separately when comparisons were made with the other tests and indices.

Photographs of the children's play constructions were scored for (a) the number of objects and (b) the area of the construction-board used. Both were reliable on a score-rescore basis, since the first involved a direct count and the second showed satisfactory interrater agreement (.88).

Clinical Ratings. In addition to the more objective scores it was desirable to have holistic, "clinical" evaluations of constriction for the projective materials. For these, raters were asked to judge the test results as they would in a clinic situation.

To provide a common definition of psychological constriction, the following description was typewritten and handed to each rater:

Webster's New International Dictionary contributes to our understanding: "Constrict: To draw together; to render narrower or smaller; to bind; cramp; to contract or cause to shrink. . . ." and: "Constricted: Bound; held tight; cramped."

In the Rorschach literature, which contains the most definitive work on it, constriction is generally agreed to be lack of spontaneity and response to external and internal stimuli; paucity of spontaneous expression and creativity; lack of response to environment and meager inner living; the substitution of conscious control for real adjustment; constraint; lack of freedom.

The raters were handed also the following instructions:

Please rate each on "constriction" as here defined, using a seven-point scale on which (1) represents most constriction and (7) represents least.

The steps of the scale were defined:

1. Extreme constriction
2. Control approaching suppression of all spontaneity
3. Preponderance of control
4. Balanced spontaneity and control
5. Preponderance of spontaneity
6. Spontaneity approaching lack of control
7. Extreme lack of control

[5]Unless specified otherwise, all correlations are tetrachoric, obtained from the diagrams of Chesire, Saffir and Thurstone [2].

[6]Rank-order correlations.

Each set of material was rated by two clinical psychologists working independently. Table 2 shows the significant and relatively high level of agreement between raters, which indicates that the global evaluations were satisfactorily stable from a score-rescore standpoint.

Results. Tables 2, 3, 4 and 5 present the results.

TABLE 2
INTERRATER AGREEMENT

Method	r	σ_r*
Rorschach	.94	.04
Drawings		
Family	.65	.11
Wish	.88	.05
Feel	.80	.10
Play	.90	.04

*Standard errors obtained from Hayes' tables (7).

TABLE 3
INTERTEST AGREEMENT OF THE MORE OBJECTIVE SCORES

	Rorschach F%		Play Area		Play N of objects	
	r	σ_r	r	σ_r	r	σ_r
Drawings						
Area						
Family	−.03	.16	.07	.16	.03	.16
Wish	−.13	.16	.20	.15	.10	.16
Feel	.00	.16	.10	.16	.02	.16
N of colors						
Family	.15	.16	−.10	.16	.27	.15
Wish	.10	.16	.06	.16	.10	.16
Feel	.15	.16	.00	.16	.05	.16
Rorschach						
F%			.10	.17	−.10	.16

TABLE 4
INTERTEST AGREEMENT OF CLINICAL JUDGMENTS

	Rorschach				Play			
	Rater I		Rater II		Rater I		Rater II	
	r	σ_r	r	σ_r	r	σ_r	r	σ_r
Drawings								
Family								
Rater I	−.01	.17	−.03	.16	−.21	.16	−.15	.16
Rater II	.01	.16	.20	.16	−.42	.17	.05	.16
Wish								
Rater I	.13	.16	.14	.16	.25	.15	.40	.15
Rater II	.06	.16	.23	.16	.18	.16	.22	.16
Feel								
Rater I	−.02	.16	−.01	.16	.00	.16	.10	.16
Rater II	.07	.16	.10	.16	.12	.16	.23	.16
Rorschach								
Rater I					.02	.16	.11	.17
Rater II					.10	.16	−.05	.16

TABLE 5
CORRELATIONS OF TEST SCORES WITH OUTSIDE CRITERIA

	Teacher's Rating						Sociometric Choice	
	Shy		Deliberate		Unsocial			
	r	σ_r	r	σ_r	r	σ_r	r	σ_r
Rorschach								
F%	—.15	.16	—.13	.16	.07	.19	—.05	.16
Rater I	.03	.19	—.06	.17	.12	.18	—.08	.19
Rater II	.27	.17	—.10	.16	.25	.10	.00	.16
Drawings								
Area								
Family	—.30	.17	.15	.19	.03	.18	—.03	.16
Wish	—.03	.17	.10	.16	—.13	.18	—.11	.16
Feel	.00	.17	.01	.16	.02	.18	.10	.16
N of colors								
Family	—.10	.18	.07	.19	.09	.19	—.02	.16
Wish	.15	.16	.04	.20	.10	.18	—.02	.16
Feel	—.24	.18	—.20	.16	.20	.18	.22	.16
Rater I								
Family	—.02	.20	—.20	.16	—.05	.18	—.22	.16
Wish	.02	.18	.05	.17	.01	.18	—.22	.17
Feel	.00	.18	.00	.19	—.18	.19	.21	.16
Rater II								
Family	—.10	.17	—.10	.16	—.15	.19	—.15	.16
Wish	.16	.17	.02	.17	—.05	.18	—.27	.15
Feel	.02	.19	—.07	.17	—.20	.19	.07	.16
Play								
Area	.00	.17	.03	.17	—.40	.20	—.20	.18
N of objects	.15	.17	—.10	.16	—.11	.18	—.17	.15
Rater I	.23	.16	—.15	.16	—.07	.18	—.13	.16
Rater II	.30	.16	.06	.17	.07	.18	—.08	.16

1. No meaningful agreement was demonstrated between tests, either on the more objective scores or on the holistic, clinical ratings of constriction. (Tables 3 and 4)

2. Neither was there meaningful agreement of any of the projective test results, objective or holistic, with the teacher's ratings or with the sociometric information. (Table 5)

3. However, the agreement between psychologists was consistently significant and relatively close. This was true whether the psychologists scored according to selected objective criteria (see pages 270 and 271) or according to holistic, "clinical" impressions. (Table 2)

DISCUSSION

Reliably scored objective indicators and "clinical" evaluations of psychological constriction failed to show consistency from test to test or from test to certain other behavioral criteria. Then these tests, as used, do not seem to measure psychological constriction in any consistent way.

Three possibilities are consistent with the data. Constriction may not be a general personality trait but may be specific to the situation rather than characteristic of the person. If it is, no other results could be obtained. The data of Hartshorne and May [4, 5, 6] were interpreted to uphold such a view of other personality factors. The disagreement among these projective tests may demonstrate the differential effect of various situational forces, rather than the uniform operation of a general personality characteristic, upon expressive behavior.

On the other hand, these experimental results do not rule out the possibility that constriction is a relatively consistent personality characteristic. They may indicate a fallacy in the assumption underlying projective techniques: if there is no causal relationship between personality factor and material produced, such results as these must occur regardless of the specificity or generality of the personality factor.

A third hypothesis is consistent with the results. Constriction may be a general trait and one that is well projected into the materials, while the methods of reading back from test results to personality are fallacious. As a more comprehensive view [1] revealed meaningful personality patterns in the Hartshorne and May data, so a reinterpretation of projective materials may demonstrate that they can reflect consistent personality characteristics in a reliable way.

This third possibility seems the best point of departure for further work. It would be presumptious to question either the continuity of personality or the possibility of developing valid projective techniques before investigating thoroughly the methods of interpreting the tests. This is especially true in view of the unsystematic derivation of methods of interpretation now in use. Also, though not easily accessible to experiment, method of interpretation may be approached more readily than either personality composition or the basic theory of projective methods.

In the search for valid methods of interpretation, clinical experience should furnish useful hypotheses concerning possible relationships between test materials and personality variables but, as the Swift study [10] and the present one indicate, such hunches should not be accepted for use until they have been verified experimentally.

Insofar as they are comparable, the results of this study are consonant with those of Swift [10] and of Magaret and Simpson [8]. Swift found no direct relationship of insecurity as measured by the Rorschach with insecurity as indicated by certain other behavior. The present study found no direct relationship of psychological constriction as measured by the Rorschach, three sets of drawings, or a play construction, with some behavior clinically considered correlative with constriction.

Magaret and Simpson compared experimentally two tests in use clinically as indices of deterioration in psychotics and found low correlations. The present study, comparing several projective tests in use clinically as indices of psychological constriction, also showed a low level of agreement among tests. Both studies emphasize that clinical acceptance is not satisfactory evidence that various tests measure the same psychological variable.

When an attempt is made to relate test results to personality, one problem is the selection of the properties of the projective materials most worth studying. Reliability may be a useful prerequisite. In the present study, investigation of some properties of children's drawings (use of color and area) suggests that it is useless to study these as indices of personality. Their random appearance precludes their direct relationship to stable personality characteristics. Such unpredictability may be found true of other properties of projective materials now in clinical use as personality indices and it may explain in part the inadequacy of present methods of evaluating personality from the materials.

Desirable also, for the systematic relating of projective materials to personality, is experimental evidence concerning what constitutes sufficient sampling of response. Retesting, perhaps a number of times, may be necessary. Failure to find systematic relationships between the various tests, in the present study, may be due to an inadequate picture of the child's reactions to each type of material, and the concept of meaningful projective response may need further definition in terms of frequency.

On the other hand, the way a person usually deals with the test material may not reveal optimally the levels of personality organization and function which these techniques may make available. The truly projective may be the uncommon response. Each person may have ready, superficial responses which do not reveal but rather mask and protect. Then only occasionally will the stimulus take the person unaware or so involve him that his response projects deeper factors. To assume the operation of "projection" simply because a "projective technique" is used may be only word magic. Rather than invariably forcing revealing responses, projective devices may only offer unusual opportunity for projection.

In that case, interest lies not in the number of retests necessary to reveal the central tendency of an individual's reactions but rather in the number necessary as a background of superficial responses against which to recognize the revelatory ones. Also, if frequency does not indicate relevance, other criteria must be set up for recognition of the deeply projective mater-

ial. An interesting relationship may be found between level of projection and degree of ego-involvement.

As for the third problem posed for this investigation, that of the reliability of interpretation, the interpsychologist agreement obtained, on both scores and ratings, lends reality to the failure to find intertest correlation, it denies the necessity of individualism in clinical judgment, and it supports the possibility of developing global indicators as well as elementaristic measures for tests.

Though it is encouraging for the development of holistic scoring schemes—and indeed for the whole of clinical psychology—to demonstrate that psychologists can give global judgments that agree, their very concurrence may be detrimental, at this stage of development, to projective tests. Clinicians' agreement may be in large part responsible for the inconsistency among clinical tools that was found in this study. Had there been less accord among clinicians in practical situations in which projective tests were used, analysis of the bases for their interpretations would have been forced. The results of this study suggest that the working level of agreement found in clinical practice may be seriously reduced when interpsychologist and intertest independence is maintained. These clinicians agreed about constriction as shown in the various productions, but their judgments could have nothing to do with a stable personality factor within the children studied. The confidence with which these clinicians agreed is a warning against accepting clinician agreement as evidence of validity.

SUMMARY

1. The Rorschach tests, three sets of drawings, and play constructions of a group of third grade children were scored reliably for psychological constriction. Scores were reliable whether obtained objectively according to selected criteria or as ratings based upon global impressions.

2. Despite the reliability with which productions could be scored, no coherence appeared among the various productions of a given child. The child rated as showing extreme constriction on one production might be rated as showing little or none on the next.

3. Neither was there meaningful agreement of any of the projective test scores or ratings with teachers' ratings on behavior clinically considered concomittant with psychological constriction, or with seclusiveness as indicated by sociometric choices.

4. The following alternative interpretations are suggested:

(a) Personality may be made up of specific traits. In that case, "constriction" would represent the central tendency of a large number of situations in which a constrictive response is possible. The data do not refute such an interpretation and are in agreement with it to the extent that reliable measures of constriction show low intercorrelations.

(b) Use of projective materials for the appraisal of a personality characteristic like constriction may involve a faulty assumption: there may be no causal connection from personality characteristic to material produced. However, the clinical success of the tests makes a blanket denial of their possibilities seem unwise.

(c) Present methods of interpreting from the production to the personality may be inadequate or fallacious, despite the agreement of clinicians in using them. Cues for proper evaluation of constriction may be overlooked for false ones by present rational methods of evaluation, and it is possible that any production must be viewed against a more complex background of information — how the individual usually responds, the occasions on which constrictive productions appear, and so on.

Received September 12, 1949.

REFERENCES

1. ALLPORT, G. W. *Personality: a psychological interpretation.* New York: Holt, 1937.
2. CHESIRE, L., SAFFIR, M., AND THURSTONE, L. L. *Computing diagrams for the tetrachoric correlation coefficient.* Chicago: University of Chicago Bookstore, 1933.
3. FRANK, L. K. *Projective methods.* Springfield, Ill.: Charles C Thomas, 1948.
4. HARTSHORNE, H., AND MAY, M. A. *Studies in deceit.* New York: Macmillan, 1928.
5. HARTSHORNE, H., AND MAY, M. A. *Studies in service and selfcontrol.* New York: Macmillan, 1929.
6. HARTSHORNE, H., MAY, M. A., AND SHUTTLEWORTH, F. K. *Studies in the organization of character.* New York: Macmillan, 1930.
7. HAYES, S. P. Tables of the standard error of

tetrachoric correlation coefficient. *Psychometrika*, 1943, **8**, 193-203.

8. MAGARET, ANN, AND SIMPSON, MARY. Two measures of deterioration in psychotic patients. *J. consult. Psychol.*, 1948, **12**, 265-270.
9. PINTNER, R. *Intelligence testing*. New York: Holt, 1930.
10. SWIFT, JOAN W. Relation of behavioral and Rorschach measures of insecurity in pre-school children. *J. clin. Psychol.*, 1945, **1**, 196-205.

OBJECTIVE RORSCHACH "SIGNS" FOR GROUPS OF NORMAL, MALADJUSTED AND NEURO-PSYCHIATRIC SUBJECTS[1]

ROY M. HAMLIN, GEORGE W. ALBEE
AND
EARL M. LELAND
VETERANS ADMINISTRATION, PITTSBURGH, PENNSYLVANIA
AND THE UNIVERSITY OF PITTSBURGH

IN a monograph entitled *An evaluation of nondirective therapy,* Muench [6] reported the results of an investigation of personality changes in a group of patients who had been counseled by this method. In his investigation, several objective criteria were used for the purpose of evaluating therapy. Among these criteria was a group of Rorschach "signs of adjustment," composed chiefly of signs defined by Hertz [4]. Through the application of these signs of adjustment to his subjects in a foretest and an aftertest, Muench presented evidence that twelve subjects "improved" on the Rorschach following nondirective therapy. The Rorschachs were scored for the presence of twenty-two objective signs, and the subjects showed a statistically significant increase in the number of satisfactory signs following therapy.

Obviously, the Rorschach may be used in other ways than by employing objective signs to evaluate therapy. An intuitive and global interpretation of the Rorschach is infinitely richer in meaning than a numerical score based upon twenty-two signs, and the validity of intuitive interpretations can certainly be tested objectively though it is not always easy to handle the complications that arise. However, the use of signs has certain advantages. It provides a simple means to permit the setting up of such an hypothesis as Muench's: The Rorschach performance of subjects shows significant changes after those subjects have undergone psychotherapy.

Muench has demonstrated that the Rorschachs of twelve subjects did change concurrently with nondirective therapy, even when the Rorschachs were condensed to a mere numerical score based upon twenty-two signs. We need not concern ourselves here with the assumption that the same Rorschachs clinically interpreted by experts would reflect the results of therapy better than the twenty-two signs, since such an assumption can itself be subjected to experimental investigation. It is sufficient to note that the results reported by Muench raised a series of questions in the minds of the writers:

First, since no control subjects[2] were used in the Muench study, might not subjects show the same Rorschach changes with a mere repetition of the test? Hamlin and Albee [3] established a control group for Muench's original group by administering a pretest and an end test without utilizing any therapy in the interim and found no significant changes between the two tests. Their results would seem to lend considerable support to Muench's conclusion that certain shifts in basic personality structure had taken place in his group concurrently with the therapeutic process.

Second, will the results of the study be con-

[1]Sponsored by the Veterans Administration and published with the approval of the Chief Medical Director. The statements and conclusions published by the authors are a result of their own study and do not necessarily reflect the opinion or policy of the Veterans Administration.

[2]More recently, Carr [2] has published results on nine therapy cases, using Rorschach signs patterned after and very similar to those used by Muench. Carr found no significant changes following therapy. Carr also refers to the unpublished study of Reader [7] who used 56 therapy cases and 15 control cases. Reader found small but statistically significant Rorschach changes apparently due to therapy, but her treatment of the Rorschach was not the same as that employed by Muench.

firmed by a replication? Muench reported on twelve subjects. The internal consistency of his results on the Rorschach, Word-Association Test, Bell Adjustment Inventory and clinical judgments added to the evidence for improvement in his subjects. Nevertheless, it is all too possible that results with twelve subjects may be due to some overlooked factor. This we hope to investigate further in the near future, probably with a group of subjects whose disorders may be of a somewhat different nature.

Third, the question arises as to whether or not the Rorschach signs utilized by Muench can be of value in estimating the adjustment of individuals from one end of the adjustment continuum to the other, assuming that such a continuum exists. In other words, will these signs differentiate between normal, slightly maladjusted, and severely maladjusted individuals to any considerable extent? Is it not possible that even though the signs of adjustment change within the same individual with therapy, the same signs do not vary between individuals at different levels of adjustment? It may be that with respect to a certain range, changes will occur within the scope of the signs used, but that outside of that range the criteria of adjustment are of an entirely different sort. For example, the highly disturbed individual may show a certain amount of "perseveration" or "contamination" in his Rorschach responses, factors which do not fall within the scope of Muench's signs. It must be remembered, of course, that Muench made use of his signs only as a rapid objective check of the Rorschach to be employed in conjunction with a number of other criteria of adjustment. Such studies as those of Miale and Harrower [5] or Buhler [1] suggest the use of other sets of signs for differentiation between levels of adjustment. However, the primary interest in the present study is to investigate the possible applicability of these particular signs with groups of subjects different from those which have been used up to this point. The hypothesis to be tested is: differences similar to those found by Muench may be shown to exist between groups of normal, mildly maladjusted, and severely maladjusted subjects when twenty "Muench signs" are used.

TABLE 1
CHRONOLOGICAL AGE AND EDUCATIONAL LEVEL OF THE SUBJECTS

Group	Mean Age	Mean Ed. Level (yrs. College Completed)
Normal	25.6	2.6
Mildly Maladjusted	25.2	2.4
Severely Maladjusted	27.6	1.6

PROCEDURE

Twenty Rorschach signs of adjustment were obtained for three groups of college-level subjects, whose age and educational levels are summarized in Table 1. The signs used by Muench consisted of eleven "personal adjustment factors" and eleven "social adjustment factors." Muench lists the factors as follows:

Personal: (1) $M+$, (2) $M - FM$, (3) W, (4) $\%F+$, undifferentiated, (5) $\%F + prim.$, (6) $H - Hd$, (7) g, (8) $(FCh - [ChF + Ch])wt$, (9) $\%A$, (10) Do, (11) P.

Social: (1) R, (2) $\%F+$, (3) FC, (4) $(FC - [CF + C])wt$, (5) C, (6) $\%VIII\text{-}X$, (7) $FCh + F(C) + Fc + FCh'$), (8) $F : FCh + F(C) + Fc + FCh'$, (9) Failures, (10) $M + FC - (C + CF)$, (11) $A + H : Ad + Hd$.

Two of the twenty-two factors were not used in the present study because their definitions were not entirely clear to the authors. The factors not used were "personal adjustment patterns" 4 and 7. Personal, social, and total adjustment scores were obtained for each group and the appropriate statistical treatment was used to discover the significance of the differences in adjustment of the groups as shown by the Rorschach signs. The use of t was deemed adequate, since it assumes normality of distribution in the underlying variables, but not necessarily in the particular sample employed.

The first group was made up of twenty students in two university Abnormal Psychology classes and was selected randomly on a nonvolunteer basis. All prospective subjects were interviewed and those who had had any previous psychiatric care were eliminated. Of those chosen, one subject was not used because of the impossibility of arranging a suitable time for testing. Those who remained will, in the succeeding discussion, be termed "normal."

The second group was composed of two subgroups. Subgroup A was made up of the twelve subjects used by Muench and Subgroup B was made up of the nineteen subjects to whom Hamlin and Albee [3] administered a pretest Rorschach. All thirty-one of these subjects were represented in the present study by their "pretest" scores, and on the basis of the descriptions given by the original investigators, were considered to be roughly comparable in terms of educational level and degree of adjustment. This group will, for the purpose of this study, be termed "mildly maladjusted."

The third group was made up of twenty patients from the VA Mental Hygiene Clinic, Pittsburgh, Pa., and was selected primarily on the basis of educational level, so that some equation was obtained with the other groups. This clinic group is considered to be "severely maladjusted."

RESULTS

No significant differences, in terms of Muench's signs, were found between any of the three groups. No significant differences can be demonstrated when the "normal" and "mildly maladjusted" groups are compared. An inspection of the means and critical ratios when the "mildly maladjusted" and "severely maladjusted" groups are compared again indicates no significant differences between the groups. Consideration of the statistics when the extreme groups are compared still yields no significant differences in the Rorschach signs of personal, social, or total adjustment. Mean numbers of satisfactory signs of adjustment for each group and critical ratios between all groups are listed in Table 2.

TABLE 2

COMPARISON OF SATISFACTORY RORSCHACH SIGNS OF ADJUSTMENT IN NORMAL (N), MILDLY MALADJUSTED (MM), AND SEVERELY MALADJUSTED (SM) GROUPS

	N	Personal	Social	Total
Normal	20	5.70	5.85	11.55
Mildly Mal.	31	5.03	6.10	11.13
Severely Mal.	20	5.15	5.55	10.70
N — MM *t*	—	1.51	.466	.686
MM — SM *t*	—	.240	1.05	.576
N — SM *t*	—	1.13	.455	.541

It can be seen that in no instance are the differences obtained statistically significant, though there is a slight general trend toward a larger number of total satisfactory signs as normality is approached.

When only six signs—those which are common to both Muench's list of twenty-two signs, and to Miale and Harrower's [5] list of nine signs—were used, it was found that the normal and mildly maladjusted groups are but slightly different, while the normal group is significantly different (at the .05 level) from the severely maladjusted group, and the difference between mildly and severely maladjusted groups approaches significance. The data are briefly summarized in Table 3.

TABLE 3

COMPARISON OF GROUPS ON SIX SIGNS

	N	Mean No. Satis. Signs	t	t_{1-3}
Normal	20	4.05		
Mildly Mal.	31	3.90	$.37_{1-2}$	2.09*
Severely Mal.	20	3.10	1.98_{2-3}	

*$P_t = .05$

ILLUSTRATIVE CASES

Two Rorschach protocols are reproduced below. These protocols were selected to illustrate the probability that the signs considered in the present study fail to reflect the personality picture actually offered in the Rorschach. Either a more exhaustive set of signs, a different set, another method of objective scoring or a clinical evaluation of the Rorschach might show highly significant differences between the groups considered in this study. It should be kept in mind, of course, that the two cases were selected as somewhat extreme illustrations.

Each of the two Rorschachs produced the same number of satisfactory Rorschach signs, i.e., eight positive signs of adjustment. The first Rorschach was given by a "normal" college student. The second Rorschach was given by a delusional schizophrenic patient with two years of college education who, when seen for examination, was roughly dressed, wore an eccentric beard and moved in a rather stiff, self-absorbed manner. He carried a Bible in his pocket. His thinking may be illustrated by his explanation of how he "correlated the lights of the fingernails—white arc of thumb—with the moon, associating the arc of the thumb with the arc of the moon." He said he had

looked at his body and didn't know whether he was dead or alive—no means of identifying himself—no way to prove that he was on this planet or born on this planet, but that when he consciously oriented the crescent of the nails to the moon, that "fixed me definitely to this planet and gave me timing. Now, this requires practice. You have to line up with the moon, check the arc until you get a natural correlation with the moon."

Case I — College "Normal"
Male — Age 23

	Free Association	*Inquiry*
I. 13″		
1.	^ A bat	1. The body and wings.
2.	^ Skull of a deer or moose. Just the horns.	2. Part of the head between the horns.
3.	> Wing of some animal.	3. The shape.
90″		
II. 15″		
1.	^ Pelvis bone of a human.	1. The way it's joined.
2.	^ Two bear heads.	2. Shape.
3.	V Chicken's head.	3. (Dd at bottom of black) The bill, comb, head.
4.	^ One of those sea crabs.	4. (Bottom red) Way it's formed — the feelers.
75″		
III. 15″		
1.	^ Two animals.	1. Woodpecker or something. The shape.
2.	^ Sea horses.	2. Head and long tail.
3.	^ Lower part of leg and foot of a deer.	3. The shape.
4.	^ Pelvis of a human.	4. The general area. Bone Structure.
5.	< Mountain with a little snow on it.	5. Shape and the white.
6.	V Head of an old man with a beard.	6. The shape.
95″		
IV. 10″		
1.	^ A bearskin rug.	1. (W) — the furry side.
2.	^ The way a turkey gobbler's comb hangs down.	2. The way it hangs. The shape.
3.	V Part of a tree.	3. A snow-covered pine with branches hanging down.
70″		
V. 7″		
1.	^ A butterfly.	1. The shape, the wings.
2.	^ A bat.	2. Shape. Wings.
3.	V Pair of tongs.	3. The shape.
4.	^ A hill.	4. The way the two mounds are shaped.
5.	< Bill of a bird.	5. It looks hooked.
70″		
VI. 17″		
1.	^ Head of a dog.	1. Eyes, nose and whiskers.
2.	^ Skin of an animal.	2. Looks like fur — with dark streak down the center.
3.	^ A butterfly.	3. The wings.
4.	^ Wings of an airplane coming at you.	4. Just one of the top whiskers on each side with the circle in the center, looks like a plane coming head on.
5.	V Waterfalls.	5. Falling down the center. Light spots are the splash of the waterfall.
85″		

Free Association	*Inquiry*
VII. 20″	
1. ^ Two heads.	1. They look like Disney heads.
2. V Skull of a long-horned deer.	2. (Bottom Dd) The shape.
3. ^ A valley in the hills.	3. The shape. The way it seems to be between two hills.
81″	
VIII. 10″	
1. > Two animals of some kind.	1. (Usual)
2. > A fish.	2. (Dd at animal's front foot). Fins and the head.
3. > Head of a dog.	3. Shape.
4. V Backbone and ribs of an animal.	4. Down the center with the three ribs coming out.
70″	
IX. 12″	
1. < Two heads of a baby.	1. The shape.
2. < A moose's head with the eye and the horns.	2. (Usual)
3. > Part of a fish.	3. The dark center part. The way it tapers and then broadens out.
4. V A volcano erupting.	4. The fumes and smoke or lava coming up. The way it rolls. The color came in as a reddish glow.
75″	
X. 15″	
1. V Sea horse.	1. The shape.
2. ^ A lion.	2. Sitting down with head up.
3. ^ Wishbone.	3. The shape.
4. ^ A ram.	4. Fluffy, wool — head lowered, running.
5. ^ Leaf of a tree.	5. The shape and color.
6. V A bridge between two hills.	6. The way it's hooked on and then arched over.
7. V The pincers of a crab.	7. The shape.
8. ^ Heart.	8. The shape.
9. ^ A St. Bernard dog.	9. Furry-like wool, lying down; the color of it, too. The shape.
143″	

Case II — Severely Maladjusted
Male — Age 32

1. 3″	
1. ^ A flying pterodactyl. Bilateral symmetry is the real impression I get.	1. I've seen pictures of flying lizards.
45″	
II.4″	
1. ^ Well, it's — uh — most striking. Dark color contrasted by the red. It does not remind me of much of anything.	1. (This seems to be more of a comment than a response).
2. ^ The heads of two animals facing each other. Calves.	2. They might be embryonic monsters. Just the head portion.
3. ^ The upper coalescence is puzzling me.	3. (As in first response)
73″	
III. 1″	
1. ^ Dehydrated Bodies. Mummified skeletons of anthropoid animals.	1. Bizarre caricature of dehydrated monkeys. Placed erect.
2. ^ Section of the spinal cord, viewed from the exact center.	2. This (center red) would concentrate that impression.
70″	
IV. 1″	
1. ^ A dry frog; the skin of it.	1. The dried pelt of a frog.
2. ^ The botanical thing — growths you see on oak trees.	2. Top center.
3. ^ Head of an animal.	3. Bottom center.
85″	

	Free Association	*Inquiry*
V. 2″		
	1. A bat. I have an impression that the portions can be pulled apart.	1. In flight.
51″		
VI. 9″		
	1. Skinned hide of an animal.	1. Despite this part (upper third).
25″		
VII. 8″		
	1. Somebody holding his thumbs up.	1. Patient demonstrates.
12″		
VIII. 7″		
	1. Colored niceness is the first impression. Can't make any other sense of it.	
	2. The upper section. The median line. Dissected skeleton of an animal.	2. From the spine to the — uh — sternum out to the rib area.
62″		
IX. 7″		
	1. Just a mystery to me. I do not get anything unless I repeat what I said about the others. I could dissect this. This might be part of a circle or sphere.	
57″		
X. 4″		
	1. A large area — an expanse — is my first impression.	1. Because these are pushed apart (pink areas).
	2. An enlargement of pretty much the same thing as I have been seeing. The median area is — uh — much larger than any of the others. A decided separation of the two. It is a consummated development.	2. Much like the rest. The properties of a sphere. Six radiuses around a periphery.
76″		

DISCUSSION

As a result of Muench's study in which definite changes in the Rorschach were demonstrated to have taken place concurrently with nondirective therapy, the authors have undertaken to investigate the following hypothesis: that differences similar to those found by Muench between foretests and aftertests in the same individuals also may be shown to exist between individuals at different levels of adjustment.

In comparing satisfactory Rorschach signs of adjustment in a "normal" group, a "mildly maladjusted" group and a "severely maladjusted" group, it has been demonstrated that no significant differences exist when Muench's list of twenty signs is used. However, when a small number of signs, common to both Muench and to Miale and Harrower are used, there appears to be a significant difference between the normal and severely maladjusted groups, and a differentiation approaching significance between mildly and severely maladjusted groups.

The results would seem to indicate that the larger set of signs used is of little value in distinguishing the degree of maladjustment throughout the distance between extremes. In view of previous findings, it may be that the signs used are relevant only to the central segment of the adjustment continuum, but do not incorporate factors that make them discriminative throughout. It may well be that the type of change found by Muench reflects the improvement in the ability to handle adequately, through particular mechanisms, the particular conflicts with which the individuals have had to deal, without an actual change in the basic personality structure.

The significance of the findings with the six Harrower signs is difficult to evaluate. At

least, it would seem to strengthen the possibility that another group of signs might differentiate in a way which those under consideration do not.

In order to illustrate the possible inefficiency of a quantitative approach in reflecting the complete personality picture shown by the Rorschach, two Rorschach protocols have been reproduced above. It seems rather apparent that a considerable discrepancy exists between the level of adjustment indicated for each by the "signs" used and that which would be indicated as a result of qualitative evaluation. It is possible that a different and more exhaustive list of signs would more closely approximate the richer interpretation obtainable through a global approach.

It may be of interest to note that the twenty normal subjects show more *Do*, no higher *%F+*, and as much pure *C* as the severely maladjusted group. The question might be raised as to whether the "normal" college students are really normal, or whether these signs are simply invalid.

The present study would appear to lend no support to the hypothesis that significant differences exist between individuals at different levels of adjustment, in terms of the specific Rorschach signs of adjustment used by Muench.

Received October 7, 1949.

REFERENCES

1. BUHLER, CHARLOTTE, BUHLER, K., AND LEFEVER, D. W. Development of the basic Rorschach score with manual of directions. Rorschach Standardization Studies No. 1. Hollywood, Calif.: Authors, 1948.
2. CARR, A. C. An evaluation of nine nondirective psychotherapy cases by means of the Rorschach. *J. consult. Psychol.*, 1949, **13**, 196-205.
3. HAMLIN, R. M. AND ALBEE, G. W. Muench's tests before and after nondirective therapy: A control group for his subjects. *J. consult. Psychol.*, 1948, **12**, 412-416.
4. HERTZ, MARGUERITE R. *Frequency tables to be used in scoring responses to the Rorschach inkblot test.* (3rd Ed.) Cleveland, Ohio: Western Reserve Univ., 1946.
5. MIALE, FLORENCE R. AND HARROWER-ERICKSON, MOLLY R. Personality structure in the psychoneuroses. *Rorschach Res. Exch.*, 1940, **4**, 71-74.
6. MUENCH, G. A. An evaluation of nondirective psychotherapy by means of the Rorschach and other indices. *Appl. Psychol. Monogr.* No. 13, Stanford, Calif.: Stanford Univ. Press, 1947.
7. READER, NATALIE. An investigation into some personality changes occurring in individuals undergoing client-centered therapy. Unpublished Ph.D. thesis, Univ. of Chicago, 1948.

RORSCHACH HUMAN AND OTHER MOVEMENT RESPONSES IN RELATION TO INTELLIGENCE[1,2]

J. E. TUCKER
THE PENNSYLVANIA STATE COLLEGE

IN formulating the theoretical framework of inkblot perception, Rorschach [17] postulated that the perception of increasing numbers of human movement responses was related to an increasing degree of intelligence. The perception of other kinds of movement and its relation to intelligence was not considered by Rorschach. Both Beck [1] and Klopfer [10] concurred with Rorschach's concept of human movement responses being highly related to the intellectual capacities of a subject. Klopfer [10, 11] further expanded movement scoring of the Rorschach test by adding two new movement score categories, animal movement and minor movement responses. The relationship between these two movement scores and intelligence has never been adequately discussed or studied in clinical groups.

Data found in studies by Bradway, *et al.* [3], Gardner [4], Harriman [6], Hertzman [7], Klopfer [9], Lindner [12], Rickers-Ovsiankina [16], Stainbrook and Siegel [18], Vernon [19] and Wishner [21] tend to confirm the hypothesis that as intelligence increases an increase in the number of human movement responses in Rorschach records may be expected. The literature on animal movement and minor movement responses in relation to intelligence is largely theoretical and inconclusive. Klopfer [8] suggests that it is a less difficult mental task to see animal shapes in movement in the inkblots and that these responses are probably not closely allied to the intellectual level of the subject. Booth [2] states that the projection of animal and inanimate object movements represents lower grades of individual development. Goldfarb [5] feels that a preponderance of animal movement responses in a Rorschach record would indicate a deficiency in intellectual creativity. Piotrowsky [14, 15] suggests that the absence of human movement responses and a preponderance of animal movement responses indicates either a low intellectual level or immature personality development. The majority of these theoretical discussions tend to consider minor and animal movement responses as more related to low intelligence or not at all related to intelligence.

STATEMENT OF THE PROBLEM

As part of a study of the Rorschach movement factor in relation to intelligence the following problem was established and studied: is there a significant difference between the correlation of summed animal and minor movement Rorschach scores with intelligence test scores and the correlation of human movement Rorschach scores with intelligence test scores in a male, veteran neurotic population?

PROCEDURE

The subjects in this study were 100 married and unmarried male adult neurotic veterans of World Wars I and II. The sample is purported to be a random sample of those male neurotic veterans in the New Jersey area who might seek hospitalization at the Lyons Veterans Administration Hospital. All cases were diagnosed as neurotic by two relatively independent judgments; that of a psychologist who diagnosed by the use of individual tests and that of a psychiatrist who evaluated each case from the viewpoint of a clinical impression gained during observations of the patient on

[1]Sponsored by the Veterans Administration and published with the approval of the Chief Medical Director. The statements and conclusions published by the author are the result of his own study and do not necessarily reflect the opinion or policy of the Veterans Administration.

[2]This article is condensed from a masters thesis conducted under the supervision of the psychology department of The Pennsylvania State College.

the ward and in diagnostic staff conferences. All cases were administered the Wechsler-Bellevue Adult Intelligence Scale, Form I, and the Rorschach Psychodiagnostic Test as routine admission procedure. The age ranged from 19 to 59 years with a mean value of 32.5. There were available 218 cases that had been tested with the Wechsler-Bellevue Test and the Rorschach test whose neurotic diagnoses had been agreed upon by both the psychologist and the psychiatrist. It was decided to use 100 of these 218 cases since this number simplified statistical procedures. Random numbers technique was used to make the final selection of the 100 cases.

All tests were rescored by the writer since the tests had originally been administered by different clinicians. From the Rorschach test all movement scores were collected. From the Wechsler test all subtest raw scores and Full Scale IQ's were gathered. The rescoring of the Rorschach movement factors was patterned after the Klopfer [10] method. Human movement (M) was scored for any whole or part human shape or near-human shape that was engaged in human-like activity. Animal movement (FM) credit was given for any animal shape or part of animal shape engaged in animal-like activity. Minor movement scores (m) were given for movement responses resulting from natural forces on inanimate objects. The movement scores were divided into two categories; human movement (M) and a summation of animal and minor movements ($FM + m$).

Scatter plots were made of each Wechsler test score distribution against each Rorschach movement score distribution. The relationship appeared linear by inspection in each case so the Pearson product-moment correlation formula was used to compute the various correlations. All of the movement scores were highly skewed toward the higher number of responses.

The mean Wechsler Full Scale IQ of the sample population was 102.45 and the standard deviation was 15.67. The normal per cent distribution of intelligence quotients indicated by Wechsler [20, p. 40] is somewhat different than that obtained from the sample population. A Chi-square test of the differences between the theoretically normal per cent distribution of IQ's and the obtained distribution of IQ's gives a value of 12.40 which approximates the significant value at the five per cent point, 12.59. The obtained distribution of IQ's is close to being significantly different from the normal distribution of IQ's as obtained in Wechsler's population.

TABLE 1

MEANS AND STANDARD DEVIATIONS OF WECHSLER SCORES AND THEIR CORRELATIONS WITH RORSCHACH MOVEMENT SCORES

Wechsler Scores	Mean	Standard Deviation	Correlations With M	Correlations With $FM + m$
Full Scale IQ's	102.45	15.67	+.262	+.350
Information	14.51	4.25	+.267	+.243
Comprehension	12.16	3.34	+.314	+.315
Digit Span	7.40	2.35	+.130	+.226
Arithmetic	7.23	2.92	+.114	+.009
Similarities	11.44	4.39	+.447	+.303
Picture Arrangement	10.28	3.50	+.214	+.063
Picture Completion	10.60	2.75	+.254	+.307
Block Design	19.63	8.55	+.288	+.267
Object Assembly	18.27	5.44	+.194	+.210
Digit Symbol	33.87	12.96	+.166	+.217

Table 1 indicates the means and standard deviations of the Wechsler Full Scale IQ's and subtest raw scores obtained from the sample population. Also displayed in Table 1 are the correlations between the two movement scores, M and $FM + m$, the Full Scale IQ's and the Wechsler subtest raw scores.

TABLE 2

OBTAINED DISTRIBUTION OF MOVEMENT SCORES, M AND $FM + m$

Number of Movement Responses	M	$FM + m$
10 and over	0	2
9	1	3
8	0	3
7	2	4
6	2	1
5	4	10
4	6	10
3	12	19
2	20	15
1	26	12
0	27	21
N, Subjects	100	100
Mean Movement Score	1.81	2.94
Standard Deviation	1.90	2.63

Table 2 gives the obtained distribution of the two movement scores, human movement (M) and the summation of animal and minor movements ($FM + m$). The extreme skew

of the movement score distributions toward the higher number of responses is evident in the data. In each case the standard deviation approximates the mean value.

DISCUSSION OF RESULTS

The number of human movement responses correlates +.262 with Full Scale Wechsler IQ's. The number of summed animal and minor movement responses correlates +.350 with Full Scale Wechsler IQ's. Using a formula suggested by Peters [13, p. 186] for testing differences in two correlations which have one array in common, the difference of .088 between these two correlations was found not to be significant. The standard error of the difference was .090 and the critical ratio, .890.

An examination of the correlations listed in Table 1 indicates that both movement scores, *M* and *FM* + *m*, tend to correlate highest with the same Wechsler subtests. For instance, the five highest ranking correlations in Table 1 for both *M* and *FM* + *m* are seen to be with the same Wechsler subtests: Information, Comprehension, Similarities, Picture Completion and Block Design. The lowest correlations of both *M* and *FM* + *m* with Wechsler subtests is with Arithmetic and other low correlations are found in the case of Object Assembly and Digit Symbol. In general, the correlations of *M* and *FM* + *m* with the Wechsler subtests have approximately the same rank order. The highest correlations in both cases are with subtests that apparently involve observation, judgment and generalizing and abstracting behavior (Information, Comprehension, Similarities, and Block Design). The lowest correlations in both cases are with tests that seem to demand rote and reproductive activity and are more related to specific learned skills (Arithmetic, Digit Symbol and Object Assembly).

While both movement scores show a positive and significant relationship to Wechsler Full Scale IQ's, the correlation values are not high enough to indicate that Rorschach movement scores could be used to predict intelligence in this particular population.

SUMMARY

Rorschach movement responses and Wechsler-Bellevue Adult Intelligence Test scores were gathered from 100 male adult neurotic veterans to determine whether there was a significant difference between *M* responses as an estimate of intelligence as compared with *FM* + *m* responses as an estimate of intelligence. It was found that human movement (*M*) scores and summed animal and minor movement (*FM* + *m*) scores correlated at the same level of significance with intelligence test scores and that the difference in the two correlations was not statistically significant. Both movement scores correlated positively and significantly with IQ's but neither correlation was high enough to indicate that movement scores could be considered predictive of intelligence. The rank order of the correlations between the two movement scores and the Wechsler subtest scores was approximately the same suggesting that the different movement scores reflect the same general areas of intelligence. Generalizations from this study to other clinical groups would not be warranted since this population was subject to several highly selective variables such as sex, military service and geographical location.

Received October 7, 1949.

REFERENCES

1. BECK, S. J. *Rorschach's test. A variety of personality pictures.* New York: Grune & Stratton, 1945. Vol. II.
2. BOOTH, G. C. Discussion on 'Some recent Rorschach problems'. *Rorschach Res. Exch.*, 1937, **2**, 48-53.
3. BRADWAY, KATHERINE, LION, E., AND CORRIGAN, H. The use of the Rorschach in a psychiatric study of promiscuous girls. *Rorschach Res. Exch.*, 1946, **10**, 105-110.
4. GARDNER, G. E. Rorschach test replies and results in 100 normal adults of average IQ. *Amer. J. Orthopsychiat.*, 1936, **6**, 32-60.
5. GOLDFARB, W. The animal symbol in the Rorschach test and an animal association test. *Rorschach Res. Exch.*, 1945, **9**, 8-22.
6. HARRIMAN, P. L. The Rorschach test applied to a group of college students. *Amer. J. Orthopsychiat.*, 1935, **5**, 116-120.
7. HERTZMAN, M. A comparison of the individual and group Rorschach test. *Rorschach Res. Exch.*, 1942, **6**, 89-108.
8. KLOPFER, B., BURCHARD, E. AND KELLEY, D. Theory and technique of Rorschach interpretation. *Rorschach Res. Exch.*, 1939, **3**, 152-194.
9. KLOPFER, B. AND DAVIDSON, H. Rorschach statistics. Part I. Mentally retarded, normal and superior adults. *Rorschach Res. Exch.*, 1938, **2**, 164-169.
10. KLOPFER, B. AND KELLEY, D. *The Rorschach*

technique. New York: World Book, 1942 and 1946.

11. KLOPFER, B. AND SENDER, SADIE. A system of refined scoring symbols. *Rorschach Res. Exch.*, 1936, **1**, 19-22.
12. LINDNER, R. M. A further contribution to the group Rorschach. *Rorschach Res. Exch.*, 1943, **7**, 7-15.
13. PETERS, C. C. AND VAN VOORHIS, W. R. *Statistical procedures and their mathematical basis*. New York: McGraw-Hill, 1940.
14. PIOTROWSKY, Z. The M, FM and m responses as indicators of changes in personality. *Rorschach Res. Exch.*, 1937, **1**, 148-156.
15. PIOTROWSKY, Z. Tentative Rorschach formulae for educational and vocational guidance in adolescence. *Rorschach Res., Exch.*, 1943, **7**, 16-27.
16. RICKERS-OVSIANKINA, MARIA. The Rorschach test as applied to normals and schizophrenic subjects. *Brit. J. med. Psychol.*, 1938, **17**, 227-257.
17. RORSCHACH, H. *Psychodiagnostics. A diagnostic test based on perception*. Berne, Switzerland: Verlag Hans Huber, 1942.
18. STAINBROOK, E. AND SIEGEL, P. S. A comparative group Rorschach study of southern Negro and white high school college students. *J. Psychol.*, 1944, **17**, 107-115.
19. VERNON, P. E. The Rorschach inkblot test, II. *Brit. J. med. Psychol.*, 1933, **13**, 179-200.
20. WECHSLER, D. *Measurement of adult intelligence*. (3rd Ed.) Baltimore: Williams and Wilkins, 1944.
21. WISHNER, J. Rorschach intellectual factors in neurotics. *Amer. J. Orthopsychiat*, 1948, **18**, 265-279.

RORSCHACH "POPULARS" AS A FUNCTION OF THE LENGTH OF RECORD

GRACE M. THOMPSON
SANTA BARBARA COUNTY SCHOOLS, SANTA BARBARA, CALIFORNIA

IN his posthumously published paper Rorschach [7] defined "vulgar" responses as those which occur at least once in every three records. The *Vulgärantwort* of the author of the inkblot test has become known in America as the "popular" response or *P*. Rorschach conceived of these responses as representing ". . .one's share in the collective or common way of sensing or perceiving things." In the aforementioned paper, which discusses a single protocol in detail, he considers such *Vulgärantworten* in relation to the total number of responses, rather than as a raw number. Interestingly, he considers the subject's ratio of 21 per cent to be low.

The current definition of the popular response has remained substantially the same as that given by Rorschach, although some differences occur among various investigators in enumerating those responses which they will accept as populars. The interpretation of this variable has been somewhat inconsistent in one respect, however, depending upon whether *P* was considered as an absolute number or as a ratio to the total number of responses or *R*.

Beck [2], for example, insists upon the absolute number of popular responses, basing his argument on two considerations: first, that the norm of 7-9 is a more stable frame of reference than a relative value, since 20 is the maximal number of *P* possible in his system, thus limiting the possible numerator of the ratio, *P*/*R*; and second, that conformity in thinking should not be measured in terms of one's *own* mental life, but in terms of that furnished by the average "healthy" individual. Klopfer [5] discusses both ratio and absolute number, yet seems to prefer the latter, insisting upon a minimum of five as prerequisite to healthy mental functioning. However, Mons [6] interprets only the ratio, stating the average to be 25-30 per cent of the total record for "normals."

The number of responses in an individual Rorschach record varies greatly within a so-called normal range. Hence it seems reasonable to suppose that one's opportunities to perceive popular responses would vary concomitantly with the length of record. If this were the case, and *P* responses were to a substantial degree a function simply of length of record, *P*% would appear to be a much more meaningful index of normality or deviation therefrom than the raw number of *P*. Thus a *P* of 7 in a record of 70*R* (10 per cent *P*) might be farther from "normal" than a *P* of 7 in a record of 28 *R* (25 per cent *P*). Similarly a *P* of 4 in a record of 16 might be far less deviant in personality terms than a *P* of 4 in a record of 30 *R*.

In order to investigate the degree of relationship between the raw numbers of *P* and *R*, the writer used the protocols of two different samples of college students from Santa Barbara College enrolled in Psychology 1A in the fall of 1946 and the summer of 1947, with *N*'s of 128 and 100, respectively. The protocols were of the group variety, Munroe's method of administration being used.

Pearson product-moment coefficients of correlation between the raw number of populars and total responses were significantly positive for both populations, yielding *r*'s of .33 and .61 for the two groups. The first group showed definite curvilinearity of relationship between *P* and *R*, for while the linear *r* was .33, one Eta was .51 (significantly nonlinear at the .01 level). The Eta of .66 for the second group was only slightly higher than the corresponding *r* of .61[1]. The relationship of *P* and *R* did not

[1]Since *P* comprises approximately one-fourth of total *R*, objections might be raised on the grounds

prove to be significantly nonlinear in the second group. The first relationship is probably the more representative, however, in showing significant curvilinearity. Cronbach [3] notes that skewness in one or both variables is a frequent cause of curvilinearity of regression, and also that "Many of the significant Rorschach scores give sharply skewed distributions for most populations." The probability that the two distributions of *R*'s here under discussion are positively skewed is evidenced by CR's of 5.25 for the first distribution and 2.36 for the second (significantly skewed at the .001 and .02 levels of confidence, respectively).

The second sample was twice administered the Rorschach during the course of the summer term, the interval between tests being four weeks. The second administration permitted the calculation of test-retest reliabilities. A Pearson product-moment coefficient of correlation of .80 was obtained as the reliability of the number of responses (*R*). This coefficient may somewhat overestimate the reliability of this variable because of the influence of memory from one test to another, an objection to test-retest reliabilities for the Rorschach to which Cronbach [3] has already called attention. That memory may have operated as a factor is attested by the statistically significant (.01 level of confidence) increase in the number of responses on the second testing, 29.36 to 36.36.

The number of popular responses (scored à la Beck) increased somewhat, but probably owing to the ceiling imposed by the maximum of twenty possible, the increase was not significant (CR of 1.50). The test-retest reliability of the popular responses was .67. As in the case of the total number of responses, this linear coefficient probably overstates the reliability somewhat. The reliabilities were, however, computed in order to get an approximation of the degree to which the two variables could maximally intercorrelate. The square root of the product of the two reliabilities, .80 and .67, yields a value of .73. This value of .73 is the maximal upper limit, according to Garrett [4], to which the two variables could intercorrelate, even assuming the relation to be perfect.

It will be remembered that the linear correlation of *P* and *R* for the second sample was .61. Corrected for attenuation this *r* rises to .84. It may well be that the "true" relation of these variables is actually higher than .84, since the test-retest reliabilities may be overstated, thus lowering the estimated relation when freed of attenuating factors. It is, therefore, clear that to a considerable extent the number of *P* is a function of *R*.

For this reason it would seem that use of *P* as a ratio would be likely to yield the more valid index of whatever personality factors may be inherent in this Rorschach variable. Altus [1] for example showed that a restricted *proportion* of populars (0-24 per cent) yielded a significant correlation with the Schizophrenia Scale of the *Minnesota Multiphasic Personality Inventory*, whereas neither *P* nor *R* singly, as an absolute value, showed such a relationship. It may be inferred that the clinical significance of *P* is masked when absolute numbers are used, since the absolute number seems to a significant degree to be merely a function of the length of record.

A footnote to this putative relationship of *P* and *R* should, however, be added: The relationship appears to be linear so long as the record contains no more than 40 *R*. After passing this number, the regression line appears to flatten, doubtless owing to the ceiling imposed by the restricted number of *P*'s, and curvilinearity of relation follows.

SUMMARY AND CONCLUSIONS

Two groups of college students, *N*'s of 128 and 100, were administered the group Rorschach, following Munroe's directions and scored by Beck's system. The second group of students was twice administered the Rorschach so that test-retest reliabilities could be computed. Following are certain of the findings pertinent to the discussion of the relationship *P* to *R*.

1. The Pearsonian relation of *P* to *R* in the second sample of 100 students was .61. The test-retest reliability of *R* was .80 for this group; for *P*, the corresponding reliability was .67.

2. The data above imply a "true" linear relation of .84, i.e. when corrected for attenua-

that the correlations are due chiefly to overlap. Although this finding would not vitiate the conclusions drawn here, such did not prove to be the case. Pearsonian coefficients of correlation for populars against *total responses minus populars* were .19 and .51 for the respective samples, with corresponding Eta's of .41 and .64.

tion. Since the test-retest reliabilities may be overstated for reasons discussed in the body of the article, it may well be that the "true" linear relation of these two variables is actually higher than .84.

3. The distribution of *R*'s for both samples was significantly skewed (.001 and .02 l.c.). The skewness of this variable may be responsible for the significant curvilinearity of regression between *P* and *R* in the first sample of 128 college students. However, in the second sample the regression was not significantly curvilinear. It seems probable that the findings for the first group of students are more nearly correct — the relation of *P* and *R* is probably curvilinear.

The general conclusion drawn from the data is that it would be advisable, if one wishes to extract clinical meaning from *P* with the associated influence of *R* to some extent ruled out, to consider *P* as a ratio and not as an absolute number. This general finding holds true for group Rorschach records with total *R* no greater than forty or fifty. Whether this finding holds true for Rorschachs individually administered is not known to the writer, but it seems a reasonable inference that it should. Simply stated, the gist of the data is: the percentage of popular responses should be more meaningful clinically than the raw number, since the latter appears to be to a very considerable extent simply a function of the length of record.

Received October 20, 1949.

References

1. Altus, W. D. Some correlates of the group Rorschach and the Schizophrenia scale of the group MMPI among two groups of "normal" college students. *J. consult. Psychol.*, 1948, **12**, 375-378.
2. Beck, S. J. *Rorschach's test: I. Basic processes.* New York: Grune & Stratton, 1944.
3. Cronbach, L. J. Statistical methods applied to Rorschach scores: a review. *Psychol. Bull.*, 1949, **46**, 391-429.
4. Garrett, H. E. *Statistics in psychology and education.* (3rd Ed.) New York: Longmans Green, 1947.
5. Klopfer, B., and Kelley, D. *The Rorschach technique.* Yonkers, N. Y.: World Book Co., 1942.
6. Mons, W. *Principles and practice of the Rorschach personality test.* Philadelphia: Lippincott, 1949.
7. Rorschach, H. *Psychodiagnostics.* Berne: Huber, 1942.

DISCRIMINATION OF THE TYPES OF SCHIZOPHRENIA BY THE WECHSLER-BELLEVUE SCALE[1]

A. EDWIN HARPER, JR.
EWING CHRISTIAN COLLEGE
ALLAHABAD, U. P., INDIA

DIFFERENTIATION among various normal and abnormal categories of subjects on the basis of subtest patterns (the interrelationships among scores on various parts of a test) has been attempted with the Stanford-Binet, Wechsler-Bellevue, and several other scales. The early work with the Binet was largely although not entirely fruitless, partly, it is claimed, because of the heterogeneous and nonspecific nature of many of the Binet subtests. The work with the Wechsler-Bellevue has produced two general conclusions: (1) that clinically differentiated groups of subjects produce different subtest patterns on the Wechsler-Bellevue; but (2) that there is a large degree of disagreement as to just what *the* pattern is for any particular group. A rough comparison between various studies may be made on the basis of the rank-orders assigned to the subtests for a particular clinical group. The rank order correlation (ρ) between the subtest ranks for schizophrenics, as obtained by Rapaport *et al.* [12] and Rabin [11] is +.18. Rapaport and Magaret [8] agree, $\rho = +.80$. Other correlations fall between these two. Reference to other articles (Webb [16], Garfield [4], Olch [9], Levine [7], Schafer [14], and Johnson [6]), and also to the reviews by Brody [1,2] Rabin [10], and Watson [15] give further evidence of the two general conclusions mentioned above. On the one hand, some investigators make extravagant claims for the diagnostic power of Wechsler-Bellevue scatter patterns; on the other hand, some authors deny completely the validity of any such pattern analysis for diagnostic purposes.

[1]This report is based on a portion of a dissertation submitted to Columbia University in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The author gratefully acknowledges the wise and fruitful guidance received from Dr. Irving Lorge, Dr. Robert L. Thorndike, Dr. Nicholas Hobbs and Dr. Percival M. Symonds.

SUBJECTS

It has often been suggested that a large measure of the disagreements among various previous investigators may be due to the small numbers of subjects used by most of them. It was therefore decided to obtain for the present study as large a sample as practicable. For this purpose all of the completed Wechsler-Bellevue test records on white schizophrenics tested in the Fairfield State Hospital in Newtown, Conn., from 1941 (the founding of their psychological department) through early 1947, were used. The sample comprised 245 subjects, all of whom had had a diagnosis of schizophrenia, unchanged for eight months or more. The hospital draws largely from a rural population. The large majority of the subjects had been tested during the first two weeks of their first admission to the hospital. Their mean age was 31.94 years, with SD of 8.83; their mean IQ was 88.25 with SD of 15.64. The group was subdivided into 90 paranoid schizophrenics, 29 hebephrenic schizophrenics, 45 catatonic schizophrenics, 35 simple schizophrenics, 29 "mixed" and "undiagnosed" schizophrenics, and 17 subjects who, because on different admissions they received different schizophrenic diagnoses, were included in the total schizophrenic population but not in any of the subgroups. These diagnoses had been made quite independently of any Wechsler-Bellevue pattern indications.

TABLE 1
MULTIPLE CORRELATIONS FOR THE SCHIZOPHRENIC SUBGROUPS

Groups Differentiated		*R* on the 21 variables	*R* on the 3 variables
90 paranoids	vs. 155 remaining schizophrenics	.640**	.420**
29 hebephrenics	vs. 216 remaining schizophrenics	.566**	.191*
45 catatonics	vs. 200 remaining schizophrenics	.406**	.299**
35 simple	vs. 210 remaining schizophrenics	.366*	
29 mixed and undiagnosed	vs. 216 remaining schizophrenics	.329	

*Significant at the 5 per cent level of confidence.
**Significant at or better than the 1 per cent level of confidence.

PROCEDURE

It was also felt that some of the lack of agreement among former investigators may have been due to the use of relatively inefficient statistical techniques. For the present study, the raw (rather than weighted) subtest scores of the Wechsler-Bellevue Scale [17] were used, it being felt that these should give more finely differentiated results. Each of the subcategories of schizophrenics was then compared with the total remaining group of schizophrenics. Groups were differentiated by means of the Discriminant Function, an adaptation by Fisher [3] of the multiple regression technique to the problem of an undistributed dichotomous criterion. This differs from the familiar multiple correlation technique only in this: Instead of using the usual product-moment correlation as the criterion, we employ instead Richardson and Stalnaker's point biserial correlation [13]:

$$r_{pbis} = \frac{M_1 - M_2}{S_T} \sqrt{pq}$$

where M_1 and M_2 are the means of the two groups, p and q the proportion of the cases falling in each group, and S_T the standard deviation of the combined groups.

In addition to the raw subtest scores of the Wechsler-Bellevue, the following scores were also included in the multiple equations: CA, education, sex (scored "0" and "1" for female and male respectively), and eight scores representing scatter (unevenness of performance) *within* each of eight of the subtests of the Wechsler-Bellevue. The latter proved to be of little value, and are not fully described here. A full explanation of the technique may be found in the published dissertation [5].

RESULTS

As has been indicated, each subgroup of schizophrenics was discriminated from the remaining schizophrenic group, a multiple correlation being calculated for each such discrimination. The multiple R was calculated on the basis of 21 variables: ten subtest scores, CA, education, sex, and the eight intrasubtest scatter scores. The three IQ's were not used, as these scores were already accounted for in the ten subtest scores. As a considerable part of these discriminations seemed to be related to gross differences in age, education, and IQ, additional multiple correlations were calculated using these three factors alone. Table 1 reports all of these multiple correlations.

Paranoid Schizophrenics. The multiple differentiation between the 90 paranoid and the remaining 155 schizophrenics, on the basis of the above-mentioned 21 variables, was R= .640. This multiple R is significant at better than the one per cent level of confidence. As the major differences between the paranoids and the rest of the schizophrenics seemed to be in age, general intelligence, and education, rather than in the subtests themselves (Tables 2 and 3), it was decided to find the value of a weighted combination of these three scores for the differentiation of these two groups. The resulting multiple R was .420. Squaring the R's for 21 and for 3 variables, the values of .41 and .18 indicate that pattern analysis of the subtests accounted for about two and a half times as much of the total variance as was accounted for by age, intelligence, and education alone. The *beta* weights of + .18, + .33, and — .11 indicate the relative contributions made by IQ, CA, and education, respectively, to this discrimination value of $R = .420$.

TABLE 2
MEAN SCORES OF THE SCHIZOPHRENIC SUBJECTS*

	Schizophrenics				
	Paranoid (90)	Hebephrenic (29)	Catatonic (45)	Simple (35)	Mixed (29)
Verbal Subtests					
1. Information	12.10	11.17	11.07	11.69	13.28
2. Comprehension	8.72	6.48	7.91	8.31	9.66
3. Digit Span	10.11	10.24	9.91	10.40	10.38
4. Arithmetic	5.60	4.59	5.29	4.31	6.31
5. Similarities	8.59	7.24	7.87	8.11	9.41
Performance Subtests					
6. Picture Arrangement	8.18	6.62	7.11	7.83	7.93
7. Picture Completion	8.13	6.66	7.49	7.57	8.97
8. Block Design	16.41	13.38	13.51	14.26	17.21
9. Object Assembly	16.79	14.31	15.33	15.14	16.03
10. Digit Symbol	29.27	21.24	27.07	29.60	32.93
Intelligence Quotients					
11. Verbal IQ	91.52	84.28	86.42	87.29	94.90
12. Performance IQ	93.93	79.14	81.93	84.49	91.21
13. Full Scale IQ	92.02	80.55	82.82	85.00	92.41
14. Chronological Age	36.40	30.00	26.87	30.09	30.45
15. Education	9.10	9.14	9.76	9.46	10.85
16. Sex†	.48	.69	.40	.49	.62

*As the eight intrasubtest scatter scores did not prove to be significant, they are not reproduced here.
†Scored: 0 = Female, 1 = Male.

From Table 2 we see that the paranoids were the highest schizophrenic group on Picture Arrangement, Object Assembly, Performance IQ, and Chronological Age. Translating roughly into Wechsler's weighted scores, however, it is seen that neither of the two subtests mentioned was more than one weighted score above the lowest mean score reported for that subtest. Table 3 indicates that Object Assembly was the only subtest which discriminated significantly between the paranoids and the remaining schizophrenics; and that Performance and Full Scale IQ, and CA all discriminated better than did any of the subtests. The high CA and IQ is consistent with what psychiatrists have long known about schizophrenic differences.

It is difficult to compare these results directly with those of any other author, as other investigators have used Wechsler's *weighted* scores, whereas this study has used the *raw* scores for the subtests. The distribution of the raw scores with the Table of Weighted Scores is not even [17, p. 228]. For example, when 20 raw scores are translated into a scale of only 18 weighted scores, some weighted scores represent only one raw score, and some have to represent two raw score units. Thus an accurate translation of mean raw scores into strictly comparable mean weighted scores would be an extremely complicated procedure, as the correlation between raw and weighted scores is never 1.00. However, rough comparisons must and will be made. Approximate weighted score equivalents of each raw score mean have been calculated, and the various mean deviations computed.

Rapaport [12] is the only investigator who has attempted a differentiation between any of the subgroups of schizophrenia, and therefore it is with his work that the following comparisons will be made. Rapaport's 26 paranoid schizophrenics had a mean age of 34, which is comparable to the mean age of 36 for the 90 paranoids of this study. These data agree with Rapaport's in that paranoids deviate positively from their own mean scores on Information, and negatively on Arithmetic and Picture Ar-

TABLE 3
POINT BISERIAL CORRELATIONS OF EACH SCHIZOPHRENIC SUBGROUP WITH THE REMAINING SCHIZOPHRENICS[a]

	Group 1[b]	Para-noid	Hebe-phrenic	Cata-tonic	Sim-ple	Mixed
	Group 2[b]	Remaining Schizophrenics				
Verbal Subtests						
1. Information		.018	—.064	—.091	—.026	.101
2. Comprehension		.055	—.021	—.074	—.018	.123
3. Digit Span		—.019	.014	—.057	.046	.030
4. Arithmetic		.085	—.102	—.004	—.154*	.141*
5. Similarities		.027	—.098	—.059	—.029	.081
Performance Subtests						
6. Picture Arrangement		.069	—.104	—.077	.003	.012
7. Picture Completion		.033	—.156*	—.075	—.054	.113
8. Block Design		.088	—.094	—.112	—.060	.079
9. Object Assembly		.151*	—.138*	—.065	—.074	.009
10. Digit Symbol		.036	—.222*	—.061	.030	.136
Intelligence Quotients						
11. Verbal IQ		.099	—.138*	—.106	—.067	.134
12. Performance IQ		.237*	—.113	—.180*	—.095	.056
13. Full Scale IQ		.185*	—.182*	—.163*	—.084	.098
14. Chronological Age		.387*	—.081	—.269*	—.085	—.063
15. Education		—.131	—.057	.045	—.010	.202*
16. Sex[c]		—.050	.133	—.103	—.020	.082

*Correlation significant. Levels of confidence as follows: 5 per cent = .138; 1 per cent = .181.
[a]As the eight intrasubtest scatter scores did not prove to be significant, they are not reproduced here.
[b]A positive correlation occurs when Group 1 is higher; a negative when Group 2 is higher.
[c]Scored: 0 = Female, 1 = Male.

rangement. These data are at variance with Rapaport's in showing deviations on Similarities, and Picture Completion and a positive deviation on Object Assembly, on all of which tests Rapaport's data indicate no deviations. There is no deviation on Digit Span, whereas Rapaport's data deviate —1.3. The correlation between the rank orders of the paranoid subtests in the two studies (Rapaport's rank orders were calculated from the data reported in the Appendix of his book) is only $\rho = +.47$.

Hebephrenic Schizophrenics. The multiple differentiation between the 29 hebephrenics and the remaining 216 schizophrenics, on the basis of the 21 variables, was $R = .566$. The discrimination on the basis of age, education, and IQ was only $R = .191$. The first of these R's is significant at better than the one per cent level, the second only at the five per cent level of confidence. Squaring the two R's, the resulting values of .32 and .04 indicate that pattern analysis accounted for eight times as much of the difference between the hebephrenics and the remaining schizophrenics as did the variables of IQ, CA, and education alone.

In the light of this, it is interesting to note (Table 2) that the hebephrenics are the lowest group on Comprehension, Similarities, all of the Performance tests, and all three IQ's. Comprehension is 3 weighted scores below the highest subtest score of its kind, and the others are 1 or 2 weighted scores below their highest competitors. Of all of these subtests, however, the hebephrenics are reliably differentiated from the *total* group made up of the rest of the schizophrenics only on Picture Completion, Object Assembly, and Digit Symbol.[2]

[2]The Hebephrenics were also reliably (5 per cent level) differentiated from the rest of the schizophrenics on the basis of intrasubtest scatter on the Digit Span test. The difference of 2.31 between Digits Forward and Digits Backward was the largest obtained for any schizophrenic subgroup. The only other intrasubtest scatter score which was significant was Object Assembly for the Simple Schizophrenics. This latter group was less "scattered" in its performance on this subtest than was any other schizophrenic subgroup.

As no other investigator has reported an attempt to differentiate hebephrenics from other schizophrenics, no comparisons can be made. These data are consistent, however, with general psychiatric opinion that the hebephrenics, the "silly minded," show considerable intellectual deterioration. What proportion of the low IQ was premorbid, and what proportion represented actual "deterioration" due to disease is not known.

Catatonic Schizophrenics. Table 1 reports the multiple discriminations between the 45 catatonics and the 200 remaining schizophrenics. On 21 variables, this was $R = .406$; on the basis of IQ, CA, and education alone this was $R = .299$. Both these figures are significant at better than the one per cent level of confidence. Comparison of the R^2's, .17 and .09, indicates that differences in intelligence, age, and education alone accounted for over half of the variance predicted by the full pattern analysis on the basis of all 21 variables. The *beta* weights for these three factors were —.14, —.24, and + .03 respectively.

The catatonics (Table 2) were the youngest group studied. Next to the hebephrenics, they had the lowest mean intelligence. On Table 3, they are seen to be differentiated on age at the one per cent level of confidence, and on both Performance and Full Scale IQ at the five per cent level of confidence. None of the subtest differences is significant.

Again, since no other investigator has reported attempts to diagnose catatonia on the basis of Wechsler-Bellevue patterns, no comparisons can be made.

Simple Schizophrenics. The 35 simple schizophrenics were differentiated from the remaining 210 schizophrenics at only the five per cent level of confidence, $R = .366$ on 21 variables. Because this figure was low, the R on 3 variables was not calculated, as it could hardly have been expected to be significant.

The simple schizophrenics stand out on Table 2 as being both the highest group on Digit Span, and the lowest on Arithmetic. But on the former they differed less than one weighted score from the most divergent group, and on the latter only two weighted scores. Table 3 indicates that they were reliably differentiated from the rest of the schizophrenics only on the basis of Arithmetic, and this at only the five per cent level of confidence.

These data agree with those on Rapaport's nine simple schizophrenics in placing Information and Object Assembly above the subtest averages, and Arithmetic, Similarities, and Picture Completion below. They disagree in that they place Digit Span and Picture Arrangement above their respective subtest means, while Rapaport's data placed them definitely below. The rank order relationship is $\rho = +.62$. Some of this lack of a high relationship may be due to differences in age between the two groups. Rapaport's hebephrenics had a mean age of 20, whereas those used for the present study had a mean age of 30. A ten-year difference in age seems inadequate, however, to account for all of the difference in pattern. But the small sample of simple schizophrenics in both of these studies raises grave doubts as to the value of even the comparisons which have been made.

Mixed and Undiagnosed Schizophrenics. The 29 mixed and undiagnosed schizophrenics were not reliably differentiated from the remaining 216 schizophrenics; $R = .329$. Though this group is the highest of all on Table 2 on Information, Comprehension, Arithmetic, Similarities, Picture Completion, Block Design, Digit Symbol, Verbal and Full Scale IQ, and education, most of these differences are not large enough to be significant. On Table 3, they are reliably differentiated from the rest of the schizophrenics only on Arithmetic (five per cent level) and education (one per cent level).

Comparing with Rapaport's 37 "Unclassified Schizophrenics," we find general agreement on age, high scores for Information and Block Design, and negative deviations for Picture Arrangement and Digit Span. The only major disagreement is on Similarities, on which Rapaport's group deviates positively from its own mean, whereas this group deviates negatively. Rapaport's Arithmetic deviation (—1.4) is considerably greater than the —0.2 of the present study. The rank difference correlation, ρ, is + .68.

SUMMARY AND CONCLUSIONS

The many disagreements as to the correct psychometric pattern for schizophrenics on the Wechsler-Bellevue test seem traceable, partly,

to the use of small populations and relatively inefficient statistical techniques in most of the studies done thus far. The present study was an attempt to measure patterns in a sample of 245 schizophrenics, by means of Fisher's discriminant function. The latter is an adaptation of the multiple regression technique to the problem of an undistributed dichotomous variable. Twenty-one scores were used, including the *raw* subtest scores, CA, education, sex, and eight measures of intrasubtest scatter. Successive groups of 90 paranoid schizophrenics, 29 hebephrenics, 45 catatonics, 35 simple schizophrenics, and 29 mixed and undiagnosed schizophrenics were compared with the total remaining schizophrenic group. Comparisons were made first on the basis of the 21 variables mentioned above, and then on the basis of only the 3 variables of IQ, CA, and education. The following conclusions were reached:

1. On the basis of the 21 variables, significant differentiations from the remaining schizophrenic group were obtained for paranoids ($R = .640$), hebephrenics ($R = .566$), and catatonics ($R = .406$) at better than one per cent level of confidence. The simple schizophrenics were differentiated at the five per cent level, ($R = .366$). The R for the mixed and undiagnosed schizophrenic group was not significant ($R = .329$).

2. For the paranoids and catatonics the full 21 variable pattern analysis accounted for only about twice as much of the variance as did the 3 variables of IQ, CA, and education alone. For the hebephrenics the combination of 21 variables was eight times as powerful as the combination of only IQ, CA, and education.

3. Rough comparisons with the investigations of Rapaport [12] were made. Twelve specific agreements and eight specific disagreements were found. Correlations for subtest rank orders were $\rho = +.47$ (paranoids), $\rho = +.62$ (simple schizophrenics), and $\rho = +.68$ (mixed and undiagnosed schizophrenics), again indicating only partial agreement.

4. The discriminant function technique has shown that pattern analysis can be as much as eight times as powerful as predictions based on the knowledge of IQ, CA, and education alone. The highest multiple correlation in the study, however, accounted for only 41 per cent of the differentiation between the paranoids and the rest of the schizophrenics (i.e., $.64^2 = .41$).

5. Although the small numbers of the subjects involved in several of the schizophrenic subcategories used makes all conclusions highly tentative, it seems reasonable to conclude that (a) pattern analysis by means of the discriminant function is a promising diagnostic technique; but that (b) the Wechsler-Bellevue Adult Intelligence Scale is not a highly adequate tool for psychiatric diagnosis. An instrument constructed more specifically for clinical diagnosis might be expected to give better results.

Received September 14, 1949.

References

1. Brody, M. B. A survey of the results of intelligence tests in psychosis. *Brit. J. med. Psychol.*, 1942, **19**, 215-257.
2. Brody, M. B. Mental testing. *J. ment. Sci.*, 1944, **90**, 127-151.
3. Fisher, R. A. The use of multiple measurements in taxonomic problems. *Ann. Eugen., Camb.*, 1937, **7**, 179-188.
4. Garfield, S. L. Wechsler-Bellevue patterns in schizophrenia. *J. consult. Psychol.*, 1948, **12**, 32-36.
5. Harper, A. E., Jr. *Differential patterns in schizophrenia on the Wechsler-Bellevue intelligence test.* Allahabad, India: The Allahabad Christian Press Ltd., 1950.
6. Johnson, L. C. Wechsler-Bellevue pattern analysis in schizophrenia. *J. consult. Psychol.*, 1949, **13**, 32-33.
7. Levine, L. S. The utility of Wechsler's patterns in the diagnosis of schizophrenia. *J. consult. Psychol.*, 1949, **13**, 28-31.
8. Magaret, Ann. Parallels in the behavior of schizophrenics, paretics and presenile non-psychotic patients. *J. abnorm. soc. Psychol.*, 1942, **37**, 511-528.
9. Olch, Doris R. Psychometric pattern of schizophrenics on the Wechsler-Bellevue test. *J. consult. Psychol.*, 1948, **12**, 127-136.
10. Rabin, A. I. Test score patterns in schizophrenics and non-psychotic states. *J. Psychol.*, 1941, **12**, 91-100.
11. Rabin, A. I. The use of the Wechsler-Bellevue scales with normal and abnormal persons. *Psychol. Bull.*, 1945, **42**, 410-422.
12. Rapaport, D., Gill, M., and Schafer, R. *Diagnostic psychological testing.* Chicago: Year Book Publishers, 1945. Vol. I.
13. Richardson, M. W., and Stalnaker, J. M. A

note on the use of biserial *r* in test research. *J. gen. Psychol.*, 1933, **8**, 463, 465.

14. SCHAFER, R. Objective and subjective aspects of testing. *J. consult. Psychol.*, 1948, **12**, 4-7.
15. WATSON. R. I. The use of the Wechsler-Bellevue scales: a supplement. *Psychol. Bull.*, 1946, **43**, 61-68.
16. WEBB, W. B. A note on the Rabin ratio. *J. consult. Psychol.*, 1947, **11**, 107-108.
17. WECHSLER, D. *The measurement of adult intelligence.* (3rd. Ed.) Baltimore: Williams & Wilkins, 1944.

THE WECHSLER-BELLEVUE PSYCHOMETRIC PATTERN IN ANXIETY NEUROSIS[1]

SAMUEL J. WARNER
VETERANS ADMINISTRATION REGIONAL OFFICE, BROOKLYN, NEW YORK

IN the literature of clinical psychology, there exist today numerous generalizations that have not as yet been subjected to adequate validation. Some of these relate to the test performance of anxiety neurotics on the Wechsler-Bellevue mental ability scale.

This paper deals with three sorts of dimensions that fall within this area: intertest variability, difference between verbal and performance test scores, and subtest pattern. Among generalizations to be found in the literature relating to these respective dimensions are the following: "test variability in neurotics is greater than that of normal subjects" [9]; "the neurotic's performance is generally lower than his verbal score" [9]; and "a significantly lower digit span and/or arithmetic score indicates anxieties interfering with attention and voluntary concentration" [8].

The general area has been treated by Lewinski [4] and Rashkis and Welsh [7], but neither study offers definitive evidence, since only 25 anxiety neurotics and 15 patients given the diagnosis of "anxiety state" were employed, respectively. Neither study offers a comparison of the test results with those of closely matched normal groups.

PURPOSE

The purpose of this report is the presentation of evidence allowing definitive treatment of the following three questions:

1. Do anxiety neurotics differ significantly from normals with respect to intertest variability?

2. Do anxiety neurotics differ significantly from normals with respect to the difference between verbal and performance subtest scores?

3. Are there any subtest patterns that allow differentiation between groups of anxiety neurotics and groups of normals?

SUBJECTS

The experimental design employed consists essentially of the comparison of the Wechsler-Bellevue data of a group of anxiety neurotics with those of a nonpsychiatric population.

Since the procurement of a control group matched with the psychiatric group for occupation (along with adequate matching with respect to the many other variables described in this section) could not be realized because of practical considerations, it was decided to employ two control groups of different occupational characteristics. This duality offers the following advantages: it tends to offset the effects of factors peculiar to one occupation, and it allows an internal check upon the degree to which we may consider each of the control groups to be representative of essentially the same things (through a consideration of the similarity of their Wechsler-Bellevue data as against those of a supposedly essentially different group, i.e., the anxiety neurotics). In this report, use will be made of the latter of these advantages through a number of comparisons and discussions.

One of the control groups consisted of thirty-three administrative and clerical male veteran employees of the Veterans Administration Regional Office, Brooklyn, New York[2]. None of these selected subjects was undergoing treatment for any reason at the time of testing, nor did any present signs of maladjustment to the examining psychologist. The second control group consisted of the thirty-two

[1]Sponsored by the Veterans Administration and published with the approval of the Chief Medical Director. The statements and conclusions published by the author are a result of his own study and do not necessarily reflect the opinion or policy of the Veterans Administration.

[2]The author is indebted to Dr. Aaron H. Cantor for the data on the veteran controls.

"well-adjusted" patrolmen described by Rapaport [6], and will henceforth be referred to as "Rapaport's controls."

The anxiety neurotics consisted of male veterans examined at the Mental Hygiene Clinic of the Veterans Administration Regional Office, Brooklyn, New York[3].

In the selection of the subjects, controls were observed regarding the following variables: psychiatric diagnosis, age, mental level, language difficulty, and sex.

Psychiatric Diagnosis. The following steps were taken in the selection of the anxiety neurotics:

1. The files of the Mental Hygiene Clinic of the VA Brooklyn Regional Office were examined for cases bearing the "screening diagnosis" of "psychoneurosis, anxiety reaction"; and 841 cases were thus found[4].

2. Of these 841 cases 144 were found to whom the complete Wechsler-Bellevue (Form I) had been administered.

3. The case histories and treatment folders of this smaller group were examined to insure that none of the following conditions existed: evidence of psychosis at any time, substantial disagreement over diagnosis, or any organic pathology. A total of sixty cases remained after use of these restricting criteria, plus those of sex and language difficulty. The psychologists' reports were in the cases histories, and where these reports indicated organicity or psychosis the cases were discarded. However, the writer did not examine any of the patients employed, and took care not to see any original test data at this stage of the study, so that the selection of cases may (to this extent) be considered independent of his personal biases regarding test interpretation. The following may also be considered factors tending to lessen the likelihood that the findings of this investigator may be accounted for in terms of the peculiarities of any single examiner: the tests were administered by a number of staff psychologists, and the psychiatric diagnoses employed were similarly made by a number of staff psychiatrists.

Age. Wechsler [9], Jastak [3], and others have given evidence of the great need for controlling this factor. For the age distribution of the sixty anxiety neurotics, the range was 19-42, the mean 28.1, and the standard deviation 5.4; the three corresponding measures of the veteran controls are respectively 22-44, 29.8, and 5.0; while those of Rapaport's controls are 25-51, 36.3, and 6.4. When the two controls are combined, the resulting range is 22-51, the mean 33.0, and the standard deviation 6.6. It follows that the anxiety neurotics and the veteran controls are closely matched with respect to age, while Rapaport's controls average about seven years older.

Mental Level. Since the magnitudes of intertest differences are functions of mental level [3, 9], matching for this variable was attempted. The range, mean, and standard deviation of the distribution of IQ's are respectively 83 - 137, 112.1, and 12.5 for the anxiety neurotics; 84 - 128, 111.3, and 9.5 for the veteran controls; 102 - 131, 118.3, and 7.1 for Rapaport's controls; and 84 - 131, 114.8, and 9.1 when the two control groups are combined. The anxiety neurotics and the veteran controls may therefore be said to be very closely matched, while Rapaport's controls average 6.5 IQ points higher.

Language Difficulty. No subjects were employed who gave evidence of any sort of language difficulty.

Sex. Since there is evidence of a relationship between sex and Wechsler-Bellevue psychometric pattern [3,9], only male subjects were employed.

METHOD

Intertest Variability. The measure of intertest variability employed is a slight modification of that suggested by Wechsler [9]: a mean weighted score was computed for every subject, from which deviations of every subtest weighted score for that subject were determined, and such deviations were summed for every subject. A measure of the intertest variability of a group is thus afforded by averaging such sums for all the individuals within a group. In order to facilitate subsequent dis-

[3]The author wishes to acknowledge his indebtedness to Dr. Bernard Locke, Chief Clinical Psychologist, Veterans Administration Regional Office, Brooklyn, N. Y., for his interest and assistance in various aspects of the research.

[4]The "screening diagnosis" was based upon a twenty to sixty-minute psychiatric examination. It was chosen over subsequent diagnoses as the latter are determined to some extent by psychological tests, and their use would therefore lessen the independence of the validating criterion. This loss of independence is believed to outweigh the advantages that might attend use of later diagnoses.

cussions, this mean score for a group will henceforth be termed the *mean variability score.*

Difference Between Verbal and Performance IQ's. Two sorts of measures are employed here. The first allows an answer to the question of whether verbal and performance IQ's, derived from each of the groups of subjects in turn, differ significantly (intragroup); while the second refers to the comparison of groups with respect to differences between the verbal and performance IQ's of individuals (intergroup). Accordingly, the first class of measures was obtained simply by averaging the verbal and performance IQ's of the members of each group in turn, noting the difference between the means for each group, and evaluating such differences for statistical significance. The second class of measures was derived by subtracting the performance IQ from the verbal IQ for every subject, determining the mean difference for all subjects within each group (with due regard for sign), noting differences between groups with respect to these means, and testing the statistical significance of such differences.

Psychometric Pattern. Bijou has defined *psychometric pattern* as "a statement . . . of the relationship existing between two or more test ratings" [1], and it is in this sense that the term will be employed in this report.

Measures of scatter were obtained by determining the difference between every subtest weighted score and each of the other ten subtest weighted scores for every subject[5].

In the comparison of groups, therefore, the mean discrepances of each of the fifty-five sets of discrepances for one group was compared with the corresponding measure of the second, and the differences were evaluated for statistical significance. It would have required considerably less labor to compare mean weighted subtest scores of the groups, but this procedure was contraindicated by the following: differences between the subtest scores of individuals tend to neutralize each other when treated in terms of mean weighted subtest scores for an entire group. This proposition will be seen to be consistent with the findings of this investigation, when the critical ratios derived from the separate use of these two methods are compared.

It is highly desirable to deal with fifty-five discrepances for every subject, rather than to employ some less laborious method such as use of a mean score or any single subtest score as reference points from which deviations of individual subtest scores might be computed. The use of a mean as a reference point reduces real differences between subtest scores, as Jastak [3] has treated at length. The use of any single subtest score as a reference point would restrict the scope of this investigation, and a sound basis for such narrowing of effort does not appear to exist.

RESULTS

Intertest Variability. The mean variability scores for the anxiety neurosis and combined control groups are respectively 17.8 and 18.5; and their difference yields a critical ratio of but 0.9. The trends indicated by the veteran and Rapaport's subgroups are essentially the same, for their corresponding mean variability scores are 18.6 and 18.3, and the critical ratio of their difference is only 0.3. We see, therefore, that the generalization cited in the introductory section to the effect that intertest variability in neurotics is greater than that of normal subjects is not substantiated by the findings, for the actual difference found is opposite in direction to that postulated, although not statistically significant.

It is of interest to note that Eysenck [2] reported that a scatter score based upon the Matrix Test "does not distinguish between normals and neurotics." He suggested that this lack of discriminatory power might be referable to the homogeneity of the five subtests employed, and that an investigation employing more heterogeneous tests was indicated. Since Eysenck specified the Wechsler-Bellevue as an example of such a heterogeneous test, the present study is in point—and it is noteworthy that these findings substantiate his own.

Difference Between Verbal and Performance IQ's. Let us first consider the intragroup differences. The mean verbal and performance IQ's were 110.4 and 112.1 respectively for the anxiety neurotics, 114.0 and 113.1 for the combined controls, 111.3 and 109.0 for the veteran controls, and 116.8 and 116.0 for Rapaport's

[5]In order to facilitate discussion, the difference between two subtest weighted scores will henceforth be termed a *discrepance.*

controls; while critical ratios relating to the difference between verbal performance IQ's within each group are 0.7, 0.5, 0.9, and 0.4 respectively.

These data, again, fail to substantiate the generalization offered in the introductory section to the effect that "the neurotic's performance is generally lower than his verbal score [9]." Although the difference offered by this investigation runs contrary to the above generalization, the present findings are consistent with those of Lewinski [4] in which, of twenty-five anxiety neurotics the performance IQ exceeds the verbal IQ in thirteen cases, while the reverse was true in eleven; as well as with the findings of Rabin [5], whose thirty-one neurotic subjects yielded a mean difference between verbal and performance IQ's of 3.03 IQ points in the direction of a greater performance score. While this reverse direction is of interest, it need hardly be said that no evidence has been offered indicating its reflection of the operation of any factor but chance.

That the absence of statistically significant differences between the part scores of the anxiety neurotics truly reflects the trend (rather than the influence of few deviate scores) is indicated by the following breakdown of frequencies: 30 cases occur in which the performance IQ exceeds the verbal IQ, 29 cases in which the reverse is true, and one tie.

The intergroup comparison analysis, according to the procedure described in the section on methodology, found that the mean of the differences between the verbal and performance IQ's of all members of the anxiety neurosis group was —0.90[6], and the corresponding measures for the combined controls, veteran controls, and Rapaport's controls were respectively 0.94, 1.91, and —0.06. Since the difference between the means of the anxiety neurotics and the combined controls yield a critical ratio of but 0.98, and the corresponding critical ratio relating to the comparison of the two control subgroups is only 0.76, the following inferences may be drawn: (a) the anxiety neurosis and the combined control groups do not differ significantly with respect to differences between the verbal and performance IQ's of their respective members, and, (b) the likelihood that this lack of a difference is due to the neutralizing effects of essential differences between the subgroups is contraindicated.

Psychometric Patterns: There exist in the literature today a large number of clinical observations bearing upon relationships between Wechsler-Bellevue scatter pattern and the neuroses in general, and anxiety neurosis in particular. While numerous, such observations arise from clinical impression rather than disciplined investigation. Because of the space limitations of this report, this section will be devoted solely to an evaluation of the two pattern relationships offered in the introductory section, and to the presentation of positive findings.

Let us first consider the contention that a relatively low Digit Span score indicates the operation of anxiety. It is to be noted (Table 1) that although the Digit Span subtest ranks last among the subtest weighted scores for the anxiety neurotics, this order is practically the same as that among the two control groups. The difference between the mean weighted score of the anxiety neurotics for this subtest and of the combined control group yields a critical ratio of but 0.6. When we turn to an inspection of the discrepances (Tables 2, 3, and 4) we find that but one of the ten critical ratios involving the Digit Span subtest (that of the discrepance between Digit Span and Picture Arrangement) is significant at the 5 per cent level. An inspection of Table 1 reveals that the size of this single critical ratio is due to the fact that the anxiety neurotics differ from the controls with respect to Picture Arrangement score rather than Digit Span score. In brief, the findings do not support the aforementioned contention. The data confirm Rashkis and Welsh [7] who found temporary inefficiency on the Digit Span subtest to be one of the signs having "least discriminatory value" regarding the "detection of anxiety."

As for the generalization that a relatively low Arithmetic score also indicates the operation of anxiety, an inspection of Table 1 reveals that the rank order of the Arithmetic subtest weighted score for the anxiety neurotics is identical with that for the combined controls. Furthermore, none of the ten critical ra-

[6]In computing this measure, the performance IQ of every subject was subtracted from his verbal IQ, so that if the former IQ happened to be smaller, the difference was positive in sign, and vice versa.

TABLE 1
MEAN WEIGHTED SUBTEST SCORES

Subtest	Veterans Controls (33)		Rapaport's Controls (32)		Combined Controls (65)		Anxiety Neurotics (60)	
	Mean Weighted Score	Rank	Mean Weighted Score	Rank	Mean Weighted Score	Rank	Mean Weighted Score	Rank
I	12.2	1	13.0	1	12.6	1	11.8	1
C	11.1	4.5	12.7	2.5	11.9	3.5	11.7	2.5
D	10.3	10	10.7	10.5	10.5	11	10.2	11
A	10.7	7.5	11.5	7	11.1	8	11.0	8
S	11.6	2	12.4	4	12.0	2	11.1	6.5
V	11.1	4.5	12.7	2.5	11.9	3.5	11.1	6.5
PA	10.5	9	10.7	10.5	10.6	9.5	11.6	4
PC	11.0	6	11.7	6	11.3	6.5	11.7	2.5
BD	10.7	7.5	11.8	5	11.3	6.5	10.9	9.5
OA	10.0	11	11.1	9	10.6	9.5	11.4	5
DS	11.5	3	11.3	8	11.4	5	10.9	9.5

TABLE 2
MEAN DISCREPANCES* OF ANXIETY NEUROTICS

	C	D	A	S	V	PA	PC	BD	OA	DS
I	—0.10	—1.63	—0.85	—0.68	—0.68	—0.23	—0.05	—0.88	—0.40	—0.87
C	—	—1.53	—0.68	—0.58	—0.48	—0.13	0.05	—0.78	—0.43	—0.87
D	—	—	0.85	0.95	0.95	1.40	1.40	0.65	1.23	0.83
A	—	—	—	0.10	0.08	0.55	0.73	—0.10	0.38	—0.07
S	—	—	—	—	0.00	0.45	0.63	—0.20	0.28	—0.18
V	—	—	—	—	—	0.48	0.63	—0.20	0.28	—0.18
PA	—	—	—	—	—	—	0.18	—0.65	—0.17	—0.63
PC	—	—	—	—	—	—	—	—0.83	—0.37	—0.82
BD	—	—	—	—	—	—	—	—	0.48	0.02
OA	—	—	—	—	—	—	—	—	—	—0.48

*In determining the sign of individual discrepances the following convention was followed: the weighted score of the subtest listed on the left side of the table was subtracted from the weighted score listed along the top of the table, so that if the former score was smaller than the latter the discrepance was positive, and vice versa. The meanings of signs in Table 2 follow accordingly.

TABLE 3
MEAN DISCREPANCES* OF COMBINED CONTROLS

	C	D	A	S	V	PA	PC	BD	OA	DS
I	—0.75	—2.12	—1.49	—0.66	—0.74	—2.08	—1.29	—1.38	—2.08	—1.26
C	—	—1.37	—0.74	0.15	0.02	—1.31	—0.54	—0.63	—1.32	—0.48
D	—	—	0.14	1.52	1.38	0.06	0.83	0.74	0.05	0.89
A	—	—	—	0.89	0.75	—0.57	0.20	0.11	—0.58	0.26
S	—	—	—	—	—0.14	—1.46	—0.68	—0.78	—1.42	—0.58
V	—	—	—	—	—	—1.32	—0.55	—0.65	—1.31	—0.46
PA	—	—	—	—	—	—	0.77	0.68	—0.02	0.83
PC	—	—	—	—	—	—	—	—0.12	—0.78	0.06
BD	—	—	—	—	—	—	—	—	—0.71	—0.02
OA	—	—	—	—	—	—	—	—	—	0.85

*See footnote to Table 2.

TABLE 4

CRITICAL RATIOS OF DIFFERENCES BETWEEN MEAN DISCREPANCES OF ANXIETY NEUROTICS AND COMBINED CONTROLS

	C	D	A	S	V	PA	PC	BD	OA	DS
I	1.7	0.8	1.4	0.1	0.1	3.6	3.2	1.0	3.3	0.8
C	....	0.2	0.1	1.9	1.4	2.2	1.3	0.3	1.7	0.8
D	....	....	1.0	0.9	0.7	2.0	0.9	0.1	1.4	0.1
A	....	....	....	1.5	1.3	1.8	0.9	0.3	1.5	0.6
S	....	....	....	....	0.4	3.5	2.9	1.1	3.0	0.8
V	....	....	....	....	....	3.8	2.9	0.9	3.2	0.6
PA	....	....	....	....	....	....	1.2	2.7	0.3	2.9
PC	....	....	....	....	....	....	....	1.5	0.8	1.9
BD	....	....	....	....	....	....	....	....	2.4	0.1
OA	....	....	....	....	....	....	....	....	....	2.7

tios of Table 4 in which the Arithmetic score enters surpasses 1.8. It may be concluded that the evidence does not support this generalization.

Turning to the *positive* findings of this study,—reference to Table 4 reveals that three differences between the mean discrepances of the anxiety neurotics and the combined controls are significant at the 5 per cent level, and that twelve such differences are significant at the 1 per cent level. When the two control groups were compared singly with the anxiety neurotics with respect to these fifteen pairs of mean discrepances, it was found that the directions of differences between the anxiety neurotics and one control group were the same in *all* fifteen pairs with the directions of differences between the anxiety neurotics and the other control group. This evidence tends to support the position that these fifteen critical ratios reflect the operation of factors other than chance.

DISCUSSION

As for the two major classes of negative findings—that neither intertest variability nor difference between verbal and performance scores served to differentiate the group of anxiety neurotics from matched normals — little more need be said than that the evidence tends to discredit the value of both these dimensions in this differential diagnosis. The sizes of groups and the natures of controls employed seem to permit definitiveness here.

With regard to the two minor negative findings — that neither relative Arithmetic nor relative Digit Span subtest scores served to differentiate the anxiety neurotics from the combined controls — the question arises as to the likelihood that differences between the control groups might have served to obliterate real differences between either of the control groups and the anxiety neurotics. In dealing with this question two sets of statistics will be employed: the mean weighted subtest scores (Table 1) and the mean discrepances (Tables 2-4).

Taking the first of these sets, two general findings may be seen in Table 1: (a) the rank orders of the subtests in the two control groups are virtually the same, and, (b) the weighted scores of Rapaport's controls are considerably higher (in general) than those of the veteran controls, which is referable to the difference in mean IQ. Specifically, however, when the Digit Span and Arithmetic subtest scores of one control group are compared with those of the other, their differences are not statistically significant. Moreover, when the Digit Span and Arithmetic subtest scores of the anxiety neurotics are compared with those of each control group in turn, the individual differences do not approach statistical significance.

Considering the second set of statistics, when the two control groups are compared with respect to each of the fifty-five discrepances, five critical ratios denoting significance at the 5 per cent level or better are found in the differences between the mean discrepances of the following pairs of subtests: 2.0 (Comprehension-Information), 2.2 (Picture Arrangement-Vocabulary), 2.6 (Digit Symbol-Vocabulary), 2.1 (Vocabulary-Information), and 3.0 (Digit Symbol-Comprehension). Several critical ratios significant at the 5 per cent level are to be expected among fifty-five measures, however, simply on the basis of chance. Such statistically significant differences need not necessarily

reflect differences with respect to the essential dimension of psychiatric "normality," since other differences, such as occupational status, exist between these two groups. These five critical ratios, moreover, cannot serve to explain the lack of substantiation offered by this investigation with respect to the two contentions under discussion, as neither the Arithmetic nor Digit Span Subtests enter into any of them.

The conclusion seems warranted that test differences between the two control groups cannot explain the absence of evidence to substantiate the contentions for the value of the Digit Span and Arithmetic Subtests as indicators of anxiety. Of course, "anxiety" is a decidedly different concept from "anxiety neurosis," and the possibility exists that no differential with respect to anxiety status actually existed between the sixty anxiety neurotics and the sixty-five controls, but the likelihood of this possibility, particularly with regard to Rapaport's "well-adjusted" controls who had been screened carefully with respect to anxiety, is remote indeed.

The positive findings are represented by the fifteen pairs of mean discrepances yielding statistically significant differences. When Table 4 is inspected for the subtests that enter most frequently into the determination of the significant ratios, it is found that the Picture Arrangement subtest enters into seven, the Object Assembly into five, and the Picture Completion, Vocabulary, Similarities, and Information subtests enter into three each. A similar inspection of Tables 2 and 3 shows those differences between subtest patterns in the experimental group and corresponding subtest patterns in the combined control group that account for the fifteen critical ratios. Such inspection reveals that among the combined controls the first three of the above subtest scores tend to be *considerably* lower than the latter three, but among the anxiety neurotics the first three subtest scores tend to be only *slightly* lower than the Information subtest scores, and to be actually higher than the Vocabulary and Similarities subtest scores. These differentials appear to account for nine of the critical ratios. Four of the remaining six critical ratios, at the 1 and 5 per cent levels of significance, are accounted for by the following: inspection of the same tables reveals that among the combined controls the Picture Arrangement and Object Assembly subtest scores tend to be considerably lower than the Block Design and Digit Symbol Subtest scores, while among the anxiety neurotics this direction is reversed, i.e., the Picture Arrangement and Object Assembly subtest scores tend to be *higher* than the Block Design and Digit Symbol scores. One of the remaining two critical ratios appears to be a function of the following: among the combined controls the Picture Arrangement subtest scores tend to be considerably lower than the Comprehension scores, while among the anxiety neurotics this direction is maintained, but is far less pronounced. The last of these critical ratios seems to follow from the tendency for the Picture Arrangement subtest scores to be but slightly higher than the Digit Span scores among the combined controls, while this direction is far more marked among the anxiety neurotics.

The following generalization is consistent with the data: the relative positions of Picture Arrangement, Object Assembly, and Picture Completion Subtest scores are not as low among the anxiety neurotics as they are among the controls.

A search for some theoretical basis to account for this differential, suggests that these three subtests appear to be more "concrete" than the Information, Comprehension, Digit Span, Similarities, Vocabulary, Block Design, and Digit Symbol subtests with which the first three subtests are compared when the fifteen statistically significant differences are encountered. If one assumes that the differential regarding concreteness is truly afforded by this twofold classification of the subtests, then the findings of this investigation permit the conclusion that anxiety neurotics do not do as poorly on concrete tasks, in comparison with performance on more abstract ones, as do normals. Conversely, the concreteness of a task is a factor tending to raise the test performance of anxiety neurotics. However, the validity of this assumption is obviously questionable, and this "theoretical basis" is offered with the realization that it is at best but a hypothesis involving poorly defined concepts.

The evidence tends to substantiate the proposition offered in the section on methodology, that the treatment of mean discrepance scores

offers a more sensitive measure of scatter than does a similar treatment of mean weighted subtest scores. When the anxiety neurotics and the combined controls were compared with respect to mean weighted scores, the highest resulting critical ratio was only 2.5 (Information), whereas inspection of Table 4 reveals that when these two groups were compared with respect to mean discrepance scores, twelve critical ratios in excess of 2.5 were found, of which the highest was 3.8.

SUMMARY

A comparative study was undertaken to test the effectiveness of the following three dimensions regarding the differentiating of anxiety neurotics from normals on the basis of the Wechsler-Bellevue mental ability scale results: intertest variability, difference between verbal and performance IQ's, and psychometric pattern. Sixty anxiety neurotics and sixty-five controls were matched with respect to age, mental level, and sex. Controls were observed regarding language difficulty, validity of psychiatric diagnosis, bias of investigator, and independence of the following two areas of data: psychiatric diagnosis and test results. The control group was composed of Rapaport's [6] sample of thirty-two "well-adjusted" patrolmen, and thirty-three veteran employees of the VA Brooklyn Regional Office. Groups were compared with respect to measures relating to these three dimensions, and differences observed were tested for statistical significance.

On the basis of the negative and positive findings of this study, and in a context consistent with its sampling limitations, the following conclusions are drawn:

1. The Wechsler-Bellevue psychometric pattern of a group of anxiety neurotics does differ significantly from that of normals in a number of respects. A hypothesis is suggested that the concreteness of a task is a factor tending to raise the performance of anxiety neurotics.

2. None of the following Wechsler-Bellevue dimensions allows the effective differentiation of a group of anxiety neurotics from a group of normals: intertest variability, difference between verbal and performance IQ's, relative Digit Span score, or relative Arithmetic score.

Received October 5, 1949.

REFERENCES

1. BIJOU, S. W. The psychometric pattern approach as an aid to clinical diagnosis — a review. *Amer. J. ment. Def.*, 1942, **46**, 354-362.
2. EYSENCK, H. J. *Dimensions of personality.* London: Routledge & Kegan Paul, 1948.
3. JASTAK, J. Problems of psychometric scatter analysis. *Psychol. Bull.*, 1949, **46**, 177-197.
4. LEWINSKI, R. J. The psychometric pattern: I. Anxiety neurosis. *J. clin. Psychol.*, 1945, **1**, 214-221.
5. RABIN, A. I. Test-score patterns in schizophrenia and non-psychotic states. *J. Psychol.*, 1941, **12**, 91-100.
6. RAPAPORT, D., GILL, M., AND SCHAFER, R. *Diagnostic psychological testing.* Chicago: The Year Book Publishers, 1945.
7. RASHKIS, H. A., AND WELSH, G. S. Detection of anxiety by use of the Wechsler Scale. *J. clin. Psychol.*, 1946, **2**, 354-357.
8. REICHARD, SUZANNE, AND SCHAFER, R. The clinical significance of scatter on the Bellevue Scale. *Bull. Menninger Clin.*, 1943, **7**, 93-98.
9. WECHSLER, D. *The measurement of adult intelligence.* (3rd Ed.) Baltimore: Williams & Wilkins, 1944.

INTELLECTUAL DETERIORATION IN THE AGED: AGREEMENT BETWEEN THE WECHSLER-BELLEVUE AND THE BABCOCK-LEVY

CHARLOTTE FOX AND JAMES E. BIRREN[1]

NATIONAL HEART INSTITUTE, NATIONAL INSTITUTE OF HEALTH, BETHESDA, MARYLAND; AND THE GERONTOLOGY SECTION, BALTIMORE CITY HOSPITALS, BALTIMORE, MARYLAND

THE purpose of this study was to investigate the agreement of two measures of intellectual loss in 60-69 year old individuals. The two measures were the Deterioration Quotient (DQ) of the Wechsler-Bellevue Intelligence Scale [10] and the Efficiency Index (EI) of the Revised Examination for the measurement of Efficiency of Mental Functioning [2], by Babcock and Levy. Both of these tests are intended to measure decline from a previous level of intellectual ability.

The measurement of the intellectual decline which accompanies age and disease is an important theoretical and practical issue. It is important to know not only the extent of such loss but also the kind and amount associated with neurological or functional disorders, and the kind and amount associated with aging apart from identifiable pathology. Obviously, such distinctions cannot be determined accurately unless our tests provide valid measures of deterioration. The present study attempts to determine whether two of our widely-used tests give comparable results. It investigates the effects of age, uncomplicated by identifiable disease which might presumably affect intellectual functions, on performance on these tests.

A comprehensive review of the literature up to 1944 in the area of mental deficit has been prepared by Hunt and Cofer [7]. A pertinent recent study is that of Copple [4], who derived from the Wechsler-Bellevue scale a score which he calls "Senescent Decline." The measure was standardized on a population of 1403 nonpsychotic white males, and consists of "a set of Beta weights, combined into a formula which differentiates younger from older subjects with the greatest possible efficiency." Copple states that the Senescent Decline value is "decidedly more effective" than the Wechsler deterioration loss method in differentiating the age groups with respect to normal age loss of intellectual functioning.

Comparisons between the Wechsler DQ and another deterioration measure, the Shipley-Hartford Conceptual Quotient, were made by Garfield and Fey [6] and Magaret and Simpson [8]. Both studies reported correlations between the two tests which were not significantly greater than zero for neuropsychiatric patients.

SUBJECTS

A total of 50 subjects was selected from institutions in the Baltimore area. These subjects met a series of criteria designed to insure a sample of "normal" persons in the sixty year decade. Since several testing sessions were devoted to each subject, it was most convenient to use hospitalized persons, as suggested by Weisenburg, Roe, and McBride [11]. Thus, subjects were used who were available for multiple testing sessions and yet were free of organic or functional conditions which would preclude their inclusion in a group of normal subjects.

The 50 subjects were selected to meet the following criteria:

1. Age. The age range was 60 to 69. This decade was chosen because persons in the sixties may show initial signs of intellectual loss and yet

[1]The authors acknowledge the technical assistance of Mr. Jack Botwinick and Mrs. Irene Staniewicz, the criticisms of Dr. Harriet Babcock, and the cooperation of the administrators and staffs of the Baltimore City Hospitals, the U. S. Marine Hospital, Baltimore, the Fort Howard Veterans Hospital, and the Levindale Hebrew Home and Infirmary.

TABLE 1
CHARACTERISTICS OF THE PRESENT POPULATION

No.	Sex	Major Diagnosis*	Major Occupation	Age	Educ.	IQ	DQ	EI
			Residents of Old People's Home					
1	M		Boilermaker	68	8	106	84	—5.8
2	M		Candy maker	69	8	98	86	—6.9
3	M		Laborer	67	5	93	42	—7.0
4	M		Laundry worker	69	8	97	116	—7.2
5	M		Truck driver	62	7	108	94	—6.2
6	M		Cigar maker	69	6	106	81	—7.5
7	M		Weaver	65	8	116	78	—5.8
8	M		Mail clerk	66	11	88	95	—7.2
9	M		Produce merchant	68	7	95	96	—4.3
10	M		Candy maker	65	7	111	100	—8.1
11	M		Laborer	65	8	103	76	—7.5
12	F		Factory worker	66	8	89	111	—1.9
13	F		Sales clerk	66	8	105	90	—5.2
14	M		Office worker	66	6	101	78	—4.2
15	F		Clerical worker	68	9	117	81	—4.5
16	M		Candy maker	63	4	92	85	—5.8
			Hospitalized Cases					
17	M	Spastic paraplegia	Farm laborer	62	6	103	116	—5.6
18	M	Hernia	Construction worker	65	7	93	85	—4.7
19	M	Arteriosclerotic ulcer of leg	Athletics instructor	66	4	95	49	—4.6
20	M	Pulmonary emphysema	Farmer	61	5	86	107	—3.4
21	F	Early multiple sclerosis	Errand girl	60	11	108	88	—6.0
22	M	Tabes dorsalis	Sales Clerk	65	8	103	82	—6.1
23	M	Glaucoma	Sheet metal worker	62	6	102	83	—5.8
24	M	Leukemia	Farmer	65	15	120	81	—4.4
25	M	Carcinoma of nose	House painter	69	8	109	95	—4.2
26	M	Duodenal ulcer	Actor	60	9	97	98	—3.7
27	M	Asthma	Stableman	68	8	96	105	—3.8
28	M	Arthritis	Manufacturer	67	13	138	117	—4.6
29	M	Fibrosis of lung following tuberculosis	Plumber	63	7	97	97	—4.0
30	M	Hernia	Carriage & auto painter	64	4	85	95	—7.0
31	F	Circulatory defect of foot	Grocery store owner	60	10	116	109	—4.1
32	M	Leg amputated	Seaman (engineer)	60	5	102	84	—6.9
33	F	Surgery of foot	Housewife	61	12	105	110	—3.5
34	M	Chronic gastritis	Seaman (cook)	62	5	89	53	—3.6
35	M	Asthma	Construction contractor	67	6	102	84	—4.1
36	M	Carcinoma of lip	Seaman (steward)	61	10	125	112	—3.4
37	M	Carcinoma of larynx	Farm laborer	62	4	94	86	—4.1
38	M	Bursitis	Ship's master	64	8	100	122	—6.3
39	M	Hyperthyroidism	Gardener	60	5	104	88	—2.2
40	M	Arteriosclerotic heart disease	Salesman	68	12	122	82	—3.7
41	M	Arteriosclerotic heart disease	Railroad laborer	63	14	97	78	—7.7
42	M	Possible old tuberculosis	Steelmill worker	62	8	86	60	—10.8
43	M	Detached retina	Seaman	65	4	102	84	—4.3
44	M	Epididymectomy	Fisherman	64	10	114	88	—4.7
45	M	Carcinoma of esophagus	Oyster fisherman & farmer	64	4	83	136	—6.6
46	M	Carcinoma of skin	Oyster fisherman	63	9	104	79	—5.2
47	M	Hernia	Baker	64	6	89	120	—5.7
48	F	Hip fracture	Housewife	63	8	87	151	—5.2
48	M	Arteriosclerotic heart disease	Auto mechanic	61	7	110	98	—4.0
50	M	Injury of brachial plexus	Painter	62	6	104	76	—4.3

*Residents of the old people's home were ambulatory and were free of diseases which require bed care.

be alert enough to cooperate readily in the test situation. The subject was asked for both his present age and his date of birth, and an age was calculated from the birthdate. If it disagreed with the age given by the subject, the age calculated from the birthdate was used [3]. The present age in years, not rounded to the nearest six months, was used. The mean age was 64.3 years.

2. Sex. Seven women and 43 men were tested.

3. Color. All subjects studied were white.

4. Native language. The subjects were persons whose native language was English: i.e., English was the first language they had learned to speak. The authors have previously shown that elderly persons with a foreign native language have significantly smaller English vocabularies than per-

sons of the same age and education whose native language is English [5].

5. Education. Only persons who had had four or more grades of formal schooling were included. A subject was given credit for the highest grade which he had attended, and need not have completed it. Professional training was given credit. The range was 4 to 15 grades of schooling, with a median of 8.1. This value is very close to the median years of school completed by the 60-69 year old white population of the United States as stated in the 1940 census: namely, 8.3 for ages 60-64 and 8.2 for ages 65-69 [12].

6. Disease. Subjects were chosen to sample mentally normal persons in the sixties. They were free of psychosis and of any neurological or other disorder which presumably might affect intellectual functioning. Cases were rejected who had had previous cerebral vascular accidents, brain disease, and senile dementia. Evidence of peripheral arteriosclerosis was not used as a criterion.

7. Vision and hearing. Subjects were screened to exclude those with defective vision or hearing, or both, which would hinder them in completing the tests.

The mean full scale Intelligence Quotient on the Wechsler-Bellevue was 100.8, σ 11.4. The IQ range was 82 to 137.

Subjects were secured from four hospitals in the Baltimore area. Sixteen subjects were residents of the Baltimore city home for the indigent aged who did not require bed care, while 10 others were new admissions. Four subjects were secured from the chronic disease wards of the Baltimore City Hospitals. The Levindale Hebrew Home and Infirmary was the source of 1 subject, while the Fort Howard Veterans Hospital provided 6 and the U. S. Marine Hospital, Baltimore, provided 13 subjects.

A summary of pertinent information on each subject is given in Table 1.

PROCEDURE

Subjects were usually given a preliminary interview in which they received a brief explanation of the purposes of the study. When information was not available from records, the interview was used to screen for age, birthdate, nativity, education, native language, occupational history, and vision and hearing. Disease history and diagnosis were obtained from the ward physician or from the patient's medical records.

Subjects were tested individually. All 50 subjects were given both the Wechsler-Bellevue and the Babcock scales, and most subjects took one or two other tests not reported here. The scales were administered in two or more sessions, depending on the number of tests given, over a period ranging from two days to three months. One session of about one to one and one-half hours was usually needed for each test, so that four or five testing periods were devoted to most subjects.

All tests were administered by the authors, who had had extensive experience in examining older persons. The Wechsler and Babcock tests were scored by the senior author and each test was checked by another psychologist.

The Wechsler Deterioration Quotients were obtained by the formula:

$$\frac{\text{Don't Hold}}{\text{Hold}} \times 100$$

The "normal deterioration loss" was added to the Deterioration Quotient [10, p. 66, Table 6A] so that normal populations at all age levels would have DQ's of 100 as a standard basis of comparison. In this case, an age allowance of 16 points, that given by Wechsler for his highest age group, 55-59, was added to each subject's DQ. Since this age correction was a constant, it had no effect upon the correlation of the DQ and EI scores. Wechsler Intelligence Quotients were calculated by means of the extrapolation formula for persons beyond age 60, as revised by Wechsler in the October, 1947 printing of his book [10, pp. 225-226].

The Babcock test was scored according to the instructions given in the test manual [2], as clarified by Dr. Babcock in an interview.

The data were analyzed as follows: (1) the mean Wechsler Deterioration Quotient (DQ) and Babcock Efficiency Index (EI) were computed, (2) the correlation between the DQ and the EI was determined, (3) interrelationships in the DQ were investigated by determining the correlation among DQ, Hold items, and Don't Hold items, and (4) interrelationships in the EI were investigated by determining the correlations among EI, Mental Efficiency score (ME), and vocabulary level (1937 Stanford-Binet vocabulary).

RESULTS

The correlation between the Wechsler Deterioration Quotient and the Babcock Efficiency Index (—.16) was not significantly greater than zero.

The mean Wechsler DQ was 91.8, σ 20.1, or a mean of 75.8 without the correction for age. This figure suggests that the present group was not, on the average, deteriorated according to Wechsler's standards. The mean Babcock EI, on the other hand, was —5.3, σ 1.6, which, according to Babcock [1], would indicate a pathological lowering of efficiency if an EI of —3.5 or lower were considered indicative. The Babcock EI does not include an allowance for age. Our uncorrected Deterioration Quotient of 75.8 agrees with Copple's figure (75.70) for his "Over 55" group, rather than with Wechsler's normal Deterioration Quotient of 84 for his 55-59 year standardization group.

Education in this group was not related to either deterioration measure. Of the correlation among the DQ, EI, age, and education, none indicated a relationship which was significantly greater than zero. Total weighted score on the Wechsler showed no relation to DQ, but was correlated significantly with education (.54).

TABLE 2

CORRELATIONS AMONG THE WECHSLER AND BABCOCK MEASURES OF DETERIORATION AND THEIR STABLE AND UNSTABLE SUBSCORES*

	Babcock EI	Weschler Hold	Wechsler Don't Hold
Wechsler DQ	—.16	—.33	.41
Babcock Vocabulary	—.24	.66	.59
Babcock ME	—.53	.74	.74

*Since the number of subjects for each correlation was 50, the standard error of r is .14 with the hypothesis that r = zero. The correlation must be .28 or larger to be significant at the 5 per cent level, or .36 or larger at the 1 per cent level.

The Babcock and Wechsler deterioration indices both compare an age-stable measure with one which changes with age. The interrelation of these age-stable and age-unstable parts of the tests offers a clue as to why the tests failed to show a close agreement. The stable indicators, the vocabulary score of the Babcock and the Hold portion of the Wechsler, correlated only .66. The unstable portions of the tests, the Mental Efficiency score and Don't Hold, correlated .74 (Table 2). Neither the stable nor the unstable portions of the two tests showed a sufficiently high correlation for individual evaluation, and it is reasonable that any derived scores based on these measures would yield even lower correlations.

The Stanford-Binet vocabulary list used in the Babcock test correlated about equally well with the Hold and Don't Hold portions of the Wechsler DQ. Similarly, the ME portion of the Babcock correlated equally well with the Hold and Don't Hold measures. The vocabulary lists of the two tests correlated .81.

The mean Senescent Decline score for this group was 85.10, σ 21.50, the approximate value which would be predicted for this age level on the basis of Copple's data [4]. The Senescent Decline score correlated .33 and .48 with the DQ and EI, respectively. Here again it is demonstrated that the derived scores which purport to measure something other than intelligence are poorly related.

DISCUSSION

According to the Wechsler DQ, the group tested is not deteriorated; according to the Babcock EI, it is deteriorated. This discrepancy might be thought to be due merely to the presence in the DQ of an allowance for normal age loss, which the EI does not include [9]. If this were the case, the DQ and EI for any given member of the group would merely differ by a constant amount, and the two tests would still show a high relationship. The negligible correlation between the DQ and EI indicates, however, that this is not the case. Though the tests are certainly not measuring the same function, the DQ and EI might both be validly measuring different aspects of intellectual loss in the elderly.

The DQ is based on only eight subtest scores, four each in the Hold and Don't Hold groups. To quote Wechsler [10], ". . .perhaps the most serious limitation to the method as a whole is the assumed very high reliability of the tests themselves. Such high order of reliability does not in fact exist."

Thus, an error of measurement in one or two subtests may introduce a marked change in the DQ. For example, if the Hold total is 20 and the Don't Hold total is 10, the DQ is 50 plus the age allowance. If 4 of the 10 Don't

Hold points were earned by repeating a total of 8 digits and on a retest the subject repeated 9 digits, his DQ would rise from 50 to 60.

Such errors of measurement are particularly likely to introduce inaccuracy at the low intelligence levels, where the few weighted score points which may change in any one subtest represent such a large proportion of the Hold or Don't Hold sum. It may be desirable to set a limiting IQ level below which the DQ cannot be considered reliable. However, it is in the low group that the question of deterioration is most frequently raised. In elderly persons, amentia may be confused with dementia.

The present population might be considered so homogeneous as to mask a true correlation between the two tests. This issue may be investigated by determining how large the variance of the test scores would have to be to yield a correlation acceptable for clinical usage. The correlation between the two tests should be at least .80 for interchangeable use of the tests in individual measurement. In the Wechsler test, for example, the standard deviation of the DQ's of the present group was 20.1. Using this value as the standard error of estimate and a correlation of .80, the required standard deviation of the population would be 33.5. A population having a standard deviation of this order is difficult to conceive, since the standard deviation of the DQ's of Wechsler's group was only about 14.8 [10, p. 66]. With a standard deviation of 33.5, it would be impossible for even a pathologically deteriorated individual to secure a score of —3 sigmas, assuming a mean of 100. While the present study does not exclude the possibility of a low positive correlation between the two tests, it does exclude the possibility that such a correlation could be high enough to allow interchangeable use of the tests for persons in the 60-69 year age range.

An extension of the present study obviously lies in applying the same tests to pathological cases matched with the present group for age, education, and language background. Until further research defines specifically the areas of validity of the two tests, their application in the older age range is of doubtful value.

SUMMARY

1. The purpose of this study was to investigate the agreement of the Wechsler-Bellevue Deterioration Quotient (DQ) and the Efficiency Index of the Babcock-Levy test (EI) as measures of intellectual loss in 60-69 year old individuals not suffering from disease which presumably might affect intellectual functioning. The two tests were administered to 50 white men and women whose native language was English and who had had four or more grades of schooling. Subjects were secured in hospitals and an old people's home.

2. The two tests showed a correlation which was not significantly greater than zero. Education had no significant effect on the deterioration scores on either test. A measure of normal age loss on the Wechsler: namely, the "Senescent Decline" devised by Copple, showed low correlations with the DQ and EI.

3. Because direct evidence of the validity of the measures is lacking for the age range of this study, it is recommended that clinical differentiation of individuals of this age level should not be attempted at present by means of the DQ and EI. It is possible that either of the measures might show a relation to an independent criterion of deterioration within a single age decade. The results of this study can offer no evidence as to the validity of the tests in detecting deterioration in young individuals who might exhibit intellectual impairment as a result of disease or trauma.

Received October 11, 1949.

REFERENCES

1. BABCOCK, HARRIET. An experiment in the measurement of mental deterioration. *Arch. Psychol., N. Y.*, 1930, **18**, No. 117.
2. BABCOCK, HARRIET, AND LEVY, LYDIA. *Test and manual of directions, the revised examination for the measurement of efficiency of mental functioning.* Chicago: C. H. Stoelting Co., 1940.
3. BIRREN, J. E., AND FOX, CHARLOTTE. Accuracy of age statements by the elderly. *J. abnorm. soc. Psychol.*, 1950, **45**, 384-387.
4. COPPLE, G. E. Senescent decline on the Wechsler-Bellevue Intelligence Scale. Unpublished Doctoral dissertation, Univ. Pittsburgh, 1948.
5. FOX, CHARLOTTE, AND BIRREN, J. E. Some factors affecting vocabulary size in later maturity: age, education, and length of institutionalization. *J. Geront.*, 1949, **4**, 19-26.
6. GARFIELD, S. L., AND FEY, W. F. A comparison of the Wechsler-Bellevue and Shipley-Hartford scales as measures of mental impairment. *J. consult. Psychol.*, 1948, **12**, 259-264.
7. HUNT, J. MCV., AND COFER, C. N. Psycho-

logical deficit. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* New York: Ronald Press, 1944. Pp. 971-1032.

8. MAGARET, ANN, AND SIMPSON, MARY M. A comparison of two measures of deterioration in psychotic patients. *J. consult. Psychol.*, 1948, **12**, 265-269.
9. RABIN, A. I. Vocabulary and efficiency levels as functions of age in the Babcock method. *J. consult. Psychol.*, 1947, **11**, 207-211.
10. WECHSLER, D. *The measurement of adult intelligence.* (3rd Ed.) Baltimore: Williams & Wilkins, 1944. October, 1947 printing.
11. WEISENBURG, T., ROE, ANNE, AND MCBRIDE, KATHARINE E. *Adult intelligence.* New York: Commonwealth Fund, 1936.
12. *Educational Characteristics of the Population of the United States, by Age: 1940.* Washington: U. S. Dept. Com., Bur. of the Census, 1943, Series P-19, No. 4.

INTELLIGENCE TEST SCORES OF NEWLY BLINDED SOLDIERS

JACOB LEVINE, AND ALAN R. BLACKBURN
TRINITY COLLEGE, HARTFORD, CONNECTICUT

THE verbal scale of the Wechsler-Bellevue Intelligence Test is well suited for use with the blind adult. It is easily administered to the blind without modification, and is well standardized for adults. In an extensive study covering 10 schools for the blind, Hayes [1] has compared the performances on the Wechsler-Bellevue of blinded and sighted students between the ages of 10 and 24. He found that on the whole there is no significant difference between the two groups. Between the ages of 10 and 17, the blind students did somewhat more poorly in the Arithmetic, Information and Similarities subtests. From 17 to 24 they caught up with the sighted students and performances were comparable (Table 1).

Whether any intelligence test which has been devised for and standardized upon a normal sighted population is wholly valid for the blind is not easily answered. A comparison with the sighted in intellectual capacity is particularly difficult in the case of individuals who were blinded at an age too early to profit from visual experiences. In these cases it would be reasonable to expect that with this reduction in sensory experiences, the opportunities for learning and associating experiences is diminished. A large number of verbal meanings and symbols are beyond their comprehension; the result is that many of the experiences which the sighted person takes for granted are impossible for the blind. Even the Wechsler-Bellevue verbal scale, despite its adaptability for the blind, and though it is administered orally, contains test items which depend upon vision for solution. For instance, the question, "If you were lost in a forest, how would you go about finding your way out?" or the question, "How tall is the average American woman?" are more difficult problems for the blinded than they are for the sighted. It is for this reason that intelligence test scores obtained by blinded

TABLE 1

COMPARISON BETWEEN W-B SUBTEST SCORES OF WECHSLER'S NORMAL GROUP, HAYES' SCHOOL BLIND, AND BLIND SOLDIERS

	Seeing (Wechsler)				School Blind (Hayes)				Blind Soldiers	
Ages	17 - 19		20 - 24		17 - 19		20 - 24		19 - 45	
No. Cases	100		120		200		64		624	
Subtest	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Information	10.5	2.88	10.4	2.86	10.9	1.95	10.5	2.76	10.7	2.96
Comprehension	9.9	3.07	10.0	2.76	11.8	2.82	11.6	2.98	11.1	3.27
Arithmetic	8.5	2.67	9.8	2.60	9.7	3.34	8.5	3.06	10.4	4.25
Digit Span	9.3	2.74	10.1	2.96	10.3	3.35	10.2	3.62	9.0	3.41
Similarities	9.5	2.25	10.2	2.06	10.6	2.93	10.7	2.91	11.0	3.30
Vocabulary	9.3	2.3	9.4	2.40	9.5	2.30	9.3	2.59	11.0	3.27
Total Wt. Score for 6 Tests	57.0		59.9		62.8		20.8		63.2	
Mean Wt. Score	9.5		10.0		10.5		10.2		10.5	
Mean IQ	101		105		108		105		108	

subjects must be interpreted with caution.

In the case of the young adult who has practically attained the full measure of his mental capacities before he was blinded, it is probable that the effects of blindness are not as devastating. Visual experiences have been integrated to form an apperceptive core upon which the individual can draw for all subsequent adjustments even without sight. Nevertheless, whether or how blindness affects intellectual functions has never been completely answered. One method of determining the effects of blindness upon intellectual functioning is the comparison of test performances of the same individuals before and after blindness.

An opportunity to compare directly the performances of a group of these young men before and after blindness presented itself at the U.S. Army Rehabilitation Center for blinded soldiers. A large segment of these men who came to Old Farms Convalescent Hospital (Special) had taken the AGCT as sighted inductees into the Army. The AGCT is a group intelligence test which consists of 3 subtests, vocabulary, arithmetic, and block counting. Following their blindness, these men were sent to Old Farms for an 18-week period of rehabilitation. As part of a battery of motor and verbal tests, the verbal scale of the Wechsler-Bellevue was administered to each trainee.

The men who came to Old Farms Convalescent Hospital were all soldiers and a few veterans who had lost their sight in the armed forces. Their blindness, the result of either injury or disease, varied from total loss of sight to restricted vision of 20/200 in one eye. As far as could be observed, the trainees represented a fair sample of the Army population. They varied in age from 18 to 49 with educational backgrounds ranging from 2 grades of grammar school to the Ph. D. degree. Geographically they came from every one of the United States and its possessions.

A comparison of AGCT test scores of 407 of the blinded cases with the Army population as a whole reveals that we are dealing with a fairly representative sample of the Army, at least as far as AGCT scores are concerned. Table 2 shows the grade distribution percentages of each sample. The Army population represents over 8,000,000 men who were processed through Reception Centers between 1940 and 1944 [6]. It will be noted that the blinded population has fewer cases at the two extremes of the AGCT scale with a heavier concentration of blind men in Army Grade IV than is found in the total Army population. On the whole, however, there is fairly close correspondence between the two samples.

TABLE 2
COMPARISONS OF GRADE DISTRIBUTIONS BETWEEN PERCENTAGE OF BLINDED AND PERCENTAGE OF TOTAL ARMY POPULATION

Army Grade	Standard Score Limits	Blinded Group Before Losing Sight N = 407 %	Total Army Population N = over 8 million %
I	130 and above	4.9	6.0
II	110 - 129	24.4	26.5
III	90 - 109	26.1	30.5
IV	60 - 89	36.7	27.7
V	59 and below	7.9	9.3
	Mean	94.4	104.37
	S.D.	23.8	23.41

The mean AGCT score for these 407 blinded cases was 94.4 with a standard deviation of 23.8. For Form Ia, which is one of the 3 variations of the AGCT, 1782 unselected cases in the Army gave a mean of 104.7 with a standard deviation of 23.41. The difference in means probably reflects the increased number of cases in the lower end of the scale in the blind group. The dispersions of test scores in the two samples are very close, as indicated by an insignificant difference between standard deviations.

On the Wechsler-Bellevue Test, the mean IQ score for 624 unselected blind cases was 106.2 with a standard deviation of 17.24. It is to be remembered that the verbal scale of the test was used, which consists of five subtests and one optional subtest: Information, Comprehension, Arithmetic, Memory Span for Digits, Similarities, and Vocabulary as the option. Subtest scores were made comparable by converting the raw scores into weighted or standard scores. By proration, the sum of the six raw scores was converted into a weighted score and then adjusted for age.

Product moment correlation between the Wechsler-Bellevue and the AGCT test scores for the 407 cases yielded a value of + .83.

This degree of relationship is comparable to most intelligence test correlations and demonstrates that there is a relatively high degree of relationship between scores on each of the tests. Reported correlations of the AGCT with other intelligence tests [6] showed only one higher correlation and that was on the Army Alpha for a population of Army men who were preselected for AGCT scores.

Correlations between the AGCT and each of the six subtests of the Wechsler-Bellevue

TABLE 3

CORRELATION BETWEEN AGCT TOTAL SCORE AND WECHSLER-BELLEVUE SUBTESTS FOR GROUP OF 407 BLINDED SOLDIERS

AGCT Vs.	R
1. Vocabulary	+.78
2. Information	+.76
3. Similarities	+.74
4. Arithmetic	+.69
5. Comprehension	+.66
6. Digits	+.51
7. Total	+.83

are shown on Table 3. It will be observed that although the correlation between AGCT and Vocabulary is high, it is not significantly higher than the one for Information which is + .76. This relationship is somewhat surprising in view of the fact that Vocabulary represents one of the three AGCT subtests. For arithmetic, which is also common to both tests, the correlation is only + .69, and is lower than the other three subtests, two of which are not represented in the AGCT.

TABLE 4

MEAN WEIGHTED SUBTEST SCORES OF 624 BLIND CASES

	Mean	Standard Deviation	Standard error of mean	Deviation from Total mean weighted score
1. Information	10.72	2.96	.12	+ .25
2. Comprehension	11.08	3.27	.13	+ .61
3. Digit Span	9.04	3.41	.14	—1.43
4. Arithmetic	10.35	4.25	.17	— .12
5. Similarities	11.03	3.30	.13	+ .56
6. Vocabulary	10.96	3.27	.13	+ .49
Mean Total	10.47			

In Table 4 are shown the mean weighted scores for our 624 blinded cases, the standard deviation and standard error of the mean of each of the six Wechsler-Bellevue subtests. The distribution of test scores by age group is shown on Table 5. The best performance for the group was on Comprehension with a mean weighted score of 11.08. Poorest performance of the group was in Digit Span where the mean weighted score was 9.04. The total mean weighted score was 10.47. It will be noted that there is very little scatter between the means; the range is only 2.04.

TABLE 5

DISTRIBUTION OF W-B TEST SCORES BY AGE GROUPS

	Age	No. Cases	Mean
	19	5	98.50
	20 - 24	154	103.96
	25 - 29	169	104.48
	30 - 34	93	102.68
	35 - 39	30	103.92
	40 - 44	7	111.20
	35 - 48	3	122.30
Mean	27.17	461	104.06

The fact that this group of blind men obtained their highest scores in tests involving reasoning and comprehension is interesting in view of Hayes' earlier findings [2] where he observed that blind children are increasingly inferior to sighted subjects in just those functions. The difference would seem to support the contention that retardations in comprehension and reasoning in blind children could be attributed to limitations in sensory experiences and consequently in opportunities to learn. Blindness which appears in maturity does not seem to affect these functions.

The greatest deviation from the total mean score is obviously the Digit Span which is more

TABLE 6

CRITICAL RATIOS OF DIFFERENCE BETWEEN EACH SUBTEST AND THE MEAN OF REMAINING SUBTESTS

	Mean	Mean of remaining subtests	Difference	Critical ratios
Information	10.72	10.42	+ .30	.86
Comprehension	11.08	10.35	+ .73	2.2
Digit Span	9.04	10.76	—1.72	4.09
Arithmetic	10.35	10.49	— .14	.37
Similarities	11.03	10.38	+ .65	1.92
Vocabulary	10.60	10.45	+ .15	.38

than double any other deviation. In Table 6 are tabulated the critical ratios between each of the subtests and the mean of the remaining five subtests. It is clear that the mean score of the Digit Span is significantly lower than the other tests, and is the only one which is below normal.

This observed inferiority of our population in the recall of digits assumes increased significance when it is found that other investigators did not obtain similar results with other blind subjects. Hayes [1] in an exhaustive study of Digit Span scores in the Stanford-Binet found that the blind are not generally inferior to the sighted children at the various age levels. Our group of subjects differ from those of Hayes in age and in duration of blindness. In addition, all of these men were soldiers, many of whom had just undergone severe emotional stress in combat.

A possible explanation for the significantly lower scores in the memory for digits may lie in the presence of emotional factors which tended to reduce attentivity. The emotional adjustments which must be made to a newly acquired disability are probably disturbing, particularly when the disability is as great as blindness. Other test data bear out this hypothesis of decreased attentivity during the initial period of adjustment to blindness. Rapaport [4] has emphasized the importance of attention in Digit Span test and considers it to be primarily a test of attention rather than one of memory. Our test results would seem to bear out this interpretation, for there appears to be no evidence for an impairment of memory in these blinded soldiers.

In Table 1 a comparison in the performance on the six Wechsler-Bellevue subtests of three different samples is presented: (1) Wechsler's original group of sighted subjects [5]; (2) Hayes' group of school blind; (3) our own group of blinded soldiers. No striking differences are evident between the three populations. In Digit Span the blinded soldiers did more poorly than the other two groups, though the difference is not as great as it is when compared with the five subtests of the same group. On the other hand, the mean performance of the blinded soldiers in arithmetic is significantly higher than the other two groups though the standard deviation is unusually high and might reflect a lower homogeneity in educational level among the soldiers than in the other groups. It is interesting to find such a difference in performance between the two blinded populations, particularly in view of the fact that the educational level of the school blind is definitely superior to the blinded soldiers. The eighth grade is the average educational level attained by the blinded soldiers.

In a recent paper, Hermann and Hackman [3] reported the results on the Wechsler-Bellevue of 1877 non-neuropsychiatric veterans and 412 neuropsychiatric veterans who had come to a VA Guidance Center. A comparison of verbal subtests scores between their sighted group and the blinded veterans shows slight but significant differences in mean scores in favor of the blinded soldiers, though there is a somewhat larger scatter in the latter group (Tables 7 and 8). If the two groups can be regarded as comparable we can find no striking

TABLE 7
WECHSLER-BELLEVUE SCORES OF SIGHTED VETERANS AT VA GUIDANCE CENTER

	No NP Disability Rating		NP Disability Rating	
Verbal Score	No. of cases	%	No. of cases	%
40-69	42	2.24	18	4.3
70-79	109	5.81	37	9.0
80-89	289	15.40	74	17.9
90-99	469	24.99	115	27.9
100-109	544	28.98	101	24.5
110-119	309	16.46	52	12.7
120-129	102	5.44	13	3.2
130-149	13	0.69	2	0.4
Total	1877		412	
Mean IQ	99.7		96.4	
SD	13.68		14.14	

TABLE 8
FREQUENCY DISTRIBUTION OF W-B IQ'S FOR 624 BLINDED CASES

Verbal W-B Score	No. Cases	%
60-70	8	1.28
70-80	33	5.29
80-90	78	12.50
90-100	114	18.27
100-110	127	20.35
110-120	111	17.79
120-130	101	16.19
130-140	48	7.69
140-150	4	0.64
Total No.	624	
Mean	106.15	
SD	17.24	

TABLE 9
W-B AND AGCT IN RELATION TO DURATION OF BLINDNESS

Duration of Blindness	2-6 months (88 cases)		7-12 months (209 cases)		13-18 months (98 cases)		19-24 months (24 cases)		25-34 months (8 cases)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
W-B	102.0	16.9	103.8	16.9	106.3	18.2	102.1	14.3	117.4	10.9
AGCT	89.8	22.4	96.5	22.8	94.6	25.7	86.2	19.7	106.3	16.1

differences in mean test scores which could be attributed to the effects of blindness. We thus observe that comparisons of our sample of blinded soldiers with other representative samples of sighted and blind subjects with regard to intelligence test scores both before and after blinding do not show our group to be at any disadvantage. We may conclude that as far as our sample is concerned no impairment of intellectual function can be observed as a consequence of blindness.

Nevertheless, it is possible that the effects of blindness upon intelligence do not appear until some time has elapsed following the blindness. It is known that with the passage of time in a large proportion of the blind population, ability to evoke voluntary visual images diminishes and often disappears. Many newly blinded soldiers reported that they no longer were able to form visual images of those even to whom they had long been close. The question at once arises, does this loss of ability reflect a more fundamental loss of intellectual powers?

A breakdown was made of 427 cases for whom the time of onset of blindness was established. The duration of blindness in these cases ranged from 2 to 34 months with an average duration of 11.0 months. A comparison of Wechsler-Bellevue and of AGCT scores for these 427 cases which have been grouped into five groups according to duration of blindness indicates no significant difference in intelligence test scores to suggest any change in intellectual capacity (Table 9). The small number of cases in the 25 to 34 months group yielded averages that were significantly higher than the others in both the Wechsler-Bellevue and the AGCT scores. It may be concluded on the basis of these comparisons that the duration of blindness within the range of 2 to 34 months does not affect intelligence test scores in the case of these newly blinded soldiers.

SUMMARY

The performances of 624 newly blinded soldiers on the verbal scale of the Wechsler-Bellevue Intelligence Test were presented and analyzed. Four hundred seven of these cases had been tested prior to their blindness as army inductees. The distributions of test scores indicate that this group represents a fair random sample of the Army population. This sample of blinded subjects was compared with other test groups, both blind and sighted, and some differences in performance were noted. The significantly poorer performance in the Digit Span test by the group was interpreted to indicate some impairment in sustained attention, possibly, symptomatic of emotional disturbances arising from adujustment to blindness. No significant difference in test performance was observed beween subjects who had been blind for different periods of time ranging from 2 to 34 months. The test results support the assumption that blindness does not affect the basic intellectual functions when it occurs in adulthood.

Received October 7, 1949.

REFERENCES

1. HAYES, S. P. Personal communication.
2. HAYES, S. P. *Contributions to the psychology of blindness.* New York: American Foundation for the Blind, 1941.
3. HERMAN, M., AND HACKMAN, R. B. Distributions of scores on the Wechsler-Bellevue Scales and the California Test of Mental Maturity at a VA guidance center. *J. appl. Psychol.*, 1948, **32**, 642-648.
4. RAPAPORT, D., GILL, M., AND SCHAFER, R. *Diagnostic psychological testing.* Chicago: Year Book Publishers, 1945. Vol. I
5. WECHSLER, D. *The measurement af adult intelligence.* (3rd Ed.) Baltimore: Williams & Wilkins, 1944.
6. Personnel Research Section, Classification & Replacement Branch. The Army General Classification Test. *Psychol. Bull.*, 1945, **42**, 760-768.

SERIAL VERSUS CONSECUTIVE ORDER ADMINISTRATION OF THE STANFORD-BINET INTELLIGENCE SCALES

ARDEN N. FRANDSEN, BETSEY R. McCULLOUGH
AND
DAVID R. STONE
UTAH STATE AGRICULTURAL COLLEGE

ON the basis of experience with the 1916 form of the Stanford-Binet Intelligence Scale, several clinicians [1, 4, 6, 11] have claimed advantages for serial order administration of Stanford-Binet test items rather than to adhere to the consecutive order in which the tests are listed in the manual. Procedures for serial testing [4, pp. 52-53, 6, pp. 87-95, 8] and more comprehensive adaptations [3] have been proposed. Since publication of the 1937 Revised Stanford-Binet Tests of Intelligence, at which time Terman and Merrill challenged the validity of the test's norms for interpreting IQ's obtained by serial testing, three experimental evaluations of serial or adaptive testing have been published. In this paper, we shall review these studies and report a fourth investigation[1].

Advantages claimed for serial testing. Among the merits claimed for grouping similar items together (such as repeating digits, comprehension questions, stringing beads, memory for sentences, etc.) and presenting them in serial order, the following have been mentioned by one or more clinicians: (a) Serial testing avoids the otherwise tedious repetition of directions [6,11]; and with careful consideration of transitions between series, they can be arranged so that "directions for one test drifts into the next" [6]. (b) Serial testing saves time, energy (fatigue), and "nerve strain" for both subject and examiner [6]. (c) A maximum of motivation and a minimum of frustration results for the subject by avoiding the long succession of failures at the end of the examination, which are inescapable in consecutive order testing [3, 5, 6]. (d) As a specific phase of better motivation, serial arrangement enhances the possibility of selecting for beginning the examination a good rapport-creating series [6]. (e) The reduced frustration, Hutt [3] has pointed out, will probably avoid the hostility otherwise aroused toward the clinic and will result in better rapport for subsequent counseling or other remedial treatment. (f) With better initial rapport and less frustration at the conclusion, Hutt has also suggested that there will be fewer chance failures (due to anxiety distractions), especially for emotionally maladjusted individuals whose frustration tolerance may be low [3]. (g) Because the examiner is oriented quickly to the limits of the subject's intelligence, a diagnosis both of general and of special abilities and disabilities is begun at the end of the first series and is progressively elaborated with each additional series [6]. (h) This constant alertness to scoring and progressive diagnosis may also help the examiner to observe qualitative "nuances of behavior and adjustment" [3]. (i) By beginning at a low level in a series, it will be more adequately insured that the subject comprehends directions for subsequent tests in the series [6]. (j) Because serial testing is equivalent, for much of the scale, to wide-range testing (finding double basals and double ceilings), which some research has indicated may be desirable [2], it will therefore result in fewer "undiscovered inversions", i.e., potential failures below the basal and successes above the ceiling [3, 6].

[1]Betsey R. McCullough. "The effects of serial testing upon the results of the Stanford-Binet Tests of Intelligence." Master of Science Thesis, Utah State Agricultural College, 1948. (Directed by Dr. Arden N. Frandsen and Dr. David R. Stone.)

As a further argument for adaptive testing, Hutt [3] points out that two assumptions on which the merit of consecutive testing is partially based are invalid: (a) Motivation is not, as is assumed, constant; it varies with variations in the order and extent of passes and failures of individuals of the same age and ability. (b) Amounts of prior practice are not, as is also assumed in consecutive order testing, equal for individuals of the same age and ability because such individuals vary widely in the range of testing required to complete an examination. And moreover, Spache [8] notes, that with shifting to find a basal age, occasional substitutions of alternatives, and the permitted alterations in order in the presence of negativism, consecutive order testing is never strictly attained in actual practice.

Terman's and Terman and Merrill's views. Terman himself anticipated in 1916 the desirability for some deviation from the consecutive order in which the tests appear in the manual. In his first manual he wrote, "The tests arranged in this revision are in the order which is usually best to follow, but one should not hesitate to depart from the order given when it seems best in a given case to do so. It is necessary to be constantly alert so that when a child shows a tendency to balk at a given type of test, such as those of memory, language, numbers, drawing, comprehension, etc., the work can be shifted to more agreeable tasks." [9, p. 130] He apparently had not, however, expected that modifications from the order in the manual would go so far as is implied in serial and adaptive testing, for in 1937 Terman and Merrill advise specifically against such practice: "The tests of each group should be given in the order in which they appear in the manual. . . .", which order "gives a minimum of monotony and also takes account of the effect of one test upon another. . . .Serial testing—that is, giving all of the tests of one type consecutively (digits, sentences, etc.)—is not permissable if the test score is to be interpreted in terms of the established norms." [10, p. 55] Consistent with Terman's original direction on order of presentation, however, they state again that it is permissable "to shift temporarily to a more agreeable task," if a child "shows resistance toward a certain type of test," [10, p. 56]

Experimental comparison of serial and consecutive order testing. The contrast between the preferred practice of several clinicians in using this very important clinical tool and the clear statement of procedure recommended by the authors of the scale indicated the need for research on the question.

Louttit, in discussing adaptive testing mentions an unpublished and uncompleted experiment done at Indiana University in which serial testing "gave results well within the limits established by Terman's data on reliability" [5, p. 95].

Spache's paper on "Serial testing with the Revised Stanford-Binet Scale" [8] is mentioned in this section only because of the implication in the title and because it presents a serial arrangement of the test items. It contributes only very indirectly to an evaluation of serial testing, since actually his subjects (70 children, ages 2 to 9 years) were given the items "in as much of a year-by-year manner as usually occurs." After giving the full scale in the usual manner, he selected the items which might be arranged serially and computed separate mental ages and IQ's on this *abbreviated* scale. The full scale and abbreviated (serial items) scale MA's and IQ's were found to be highly similar. The mean IQ's were 122.8 and 123.0, respectively; between which the critical ratio of the difference was .31. The correlation between full scale and abbreviated scale IQ's was .946. And the IQ difference for single individuals ranged from +10 to —11 points. Spache himself admitted that, "Of course, this evidence does not prove that the use of these particular tests would yield results identical with those obtained by the entire scale, when the items are *given* in serial arrangement" [8, p. 84].

Hutt proposed in his experiment [3] to test the hypothesis that an adaptation of the Stanford-Binet "which does not affect the medium age-placement of the items, does not disturb the norms of the test", and moreover with an adaptation which eliminates the frustration from meeting a long sequence of failures at the end, the rest "will yield higher and more valid ratings for maladjusted subjects." He tested 630 individuals from kindergarten to grade 9B approximately alternately as they came to the clinic either by consecutive ($N =$

290) or adaptive ($N = 340$) methods. The children in both groups were similar in age, grade, and in S-B vocabulary score. The examiners were guided in testing the experimental group by the following principles of adaptive testing: (a) Testing was begun at a level low enough (4 years below anticipated mental level) to insure initial success. (b) Items requiring concentration, speed, prolonged and involved directions were avoided as initial tests. (c) Serial tests were administered as early as possible in order to establish basal and ceiling levels. (d) "Easy" and "hard" items (from near basal and near ceiling) were presented alternately. (e) Ordinarily, items outside the range of the basal and ceiling limits were not administered. The results confirmed both aspects of Hutt's hypothesis. The medians, which would be unaffected by cases outside the middle range (including those in the consecutive testing group whose IQ's might be lowered by frustration), were found to be practically identical. The mean IQ's of three "adaptive" subgroups were consistently slightly higher (from 1.8 to 3.8 points) than those of the corresponding "consecutive" subgroups. Separately these differences were not statistically significant, but the overall difference between the methods was significant at the 5 per cent level. For testing the second aspect of his hypothesis, Hutt selected (on the basis of "comprehensive clinical data") 24 matched pairs of maladjusted children and 33 matched pairs of well-adjusted children, and compared in each group the IQ's from adaptive and consecutive testing. As might have been expected from the results on the larger group, for the well adjusted children the mean IQ's (109.6 and 110.4) for consecutive and adaptive testing did not differ significantly. For the poorly adjusted group, however, adaptive testing resulted in a higher mean of IQ's (102.7 compared to 91.7 for consecutive testing); and these higher IQ's were, according to Hutt's hypothesis, interpreted as being more valid measures of the abilities of the maladjusted children.

THE EXPERIMENT

The junior author of this paper designed her experiment [7] to test even more specifically than Hutt had done the hypothesis that serial presentation of the Revised Stanford-Binet items (in so far as they can be arranged for such presentation) will not invalidate the use of Terman and Merrill's norms.

For Form L (both forms were similiarly arranged), the following items were arranged for serial presentation: (1) Three Hole Form Board: non rotated, Yr. II; rotated, Yr. II-6. (2) Block Building: tower, Yr. II; bridge, Yr. III. (3) Repeating digits: 2 at Yr. II-6; 3 at Yr. III; 4 at Yr. IV-6; 5 at Yr. VII; 6 at Yr. X; 8 at Yr. SAII; 9 at Yr. SAIII. (4) Stringing beads: quantity only, Yr. III; Copy From memory I, Year VI; Copy From Memory II, Yr. XIII. (5) Copying figures: Circle, Yr. III; Square, Yr. V; Diamond, Yr. VII. (6) Response to Pictures: I, Yr. III-6; II, Yr. XII. (7) Comprehension: I, Yr. III-6; II, Yr. IV; III, Yr. VII; IV, Yr. VIII. (8) Opposite Analogies: I, Yrs. 1V-6 and VII; II, Yr. SAIII. (9) Definitions: Vocabulary, Yrs. VI to SAIII; Abstract Words I, Yr. XI; Abstract Words II, Yrs. XII and XIV. (10) Memory for Sentences: II, Yr. V; III, Yr. VIII; IV, Yr. XI; V, Yr. A.A. (11) Number: Counting 4 objects, Yr. V; number concepts, Yr. VI. (12) Picture Absurdities: I, Yr. VII; II, Yr. X; III, Yr. XIV. (12) Similarities: 2 things, Yr. VII; 3 things, Yr. XI; Essential Similarities, Yr. S.A.I. (13) Verbal Absurdities: 1, Yr. VIII; II, Yr. IX and XII; III, Yr. XI. (14) Paper Cutting: I, Yrs. IX and XII; III, Yr. SAIII. (15) Repeating digits reversed: 4 at Yr. IX; 5 at Yr. XI; 6 at Yr. SAI. (16) Finding reasons: I, Yr. X; II, Yr. SAII. (17) Problems: Problem situation, Yr. XI, Problems of Fact, Yr. XIII. (18) Orientation: Direction I, Yr. XIV; Direction II, Yr. SAIII. (19) Proverbs: I, Yr. AA; II, Yr. SAII.

Out of a total of 122 items (including items scored in multiple positions) in Form L, 68 items were grouped serially. An additional 7 separate tests are scored in 17 *serial* positions. The remaining 37 separate items were left ungrouped.

In serial administration of the scale, the testing was started in the usual manner, one year level below the chronological age of the subject. As long as one of the items which had been arranged for serial presentation did not appear, the test was administered in the standard sequence as prescribed in the manual. When an item belonging to a serially arranged group occurred, the others in its series, from that year level up to the first failure, were given. Then the examiner came back to the test item at the tentatively assumed and progressively adjusted basal level following in sequence the first test given in series. If the subject failed an item, serial or nonserial, at the

tentatively assumed basal age, the examiner turned back to the preceding year level until a level was found on which all items were passed. Similarly, the test was continued upward, with both serial and nonserial items, until a level was reached where all items were failed.

Forty-seven subjects, ranging in age from 5 to 18 years, were administered both Form L and Form M of the Revised Stanford-Binet Scale, one form being given in serial arrangement and one form being administered in the consecutive order in which the items appear in the manual. To control the possible effects of practice both from one form to another and from serial arrangement to consecutive arrangement, the children were divided into four equal groups (three of 12 each and one of 11), and tested according to the schedule presented in Table 1.

TABLE 1

SCHEDULE OF TESTING TO CONTROL PRACTICE EFFECTS BOTH FROM FORM L TO M AND FROM SERIAL TO CONSECUTIVE

Group	First Test	Retest
1	Form L, Consecutive	Form M, Serial
2	Form L, Serial	Form M, Consecutive
3	Form M, Consecutive	Form L, Serial
4	Form M, Serial	Form L, Consecutive

Results. The overall data on IQ's obtained on the 47 subjects tested by both serial and consecutive order procedures are presented in Table 2.

TABLE 2

INTELLIGENCE QUOTIENTS OBTAINED FROM TESTING 47 CHILDREN WITH BOTH SERIAL AND CONSECUTIVE ARRANGEMENT OF THE REVISED STANFORD-BINET INTELLIGENCE SCALES

Arrangement	N	Mean	SD	M. Dif.	t*	r
Serial	47	109.00	19.35	.85	.81	.93
Consecutive	47	108.15	18.60			

*By formula for correlated scores.

As Table 2 indicates, the results from the two methods of testing are found to be practically identical. To justify rejecting the hypothesis that the two means (109.00 and 108.15) could occur as random samples from administration of the *same* test (with practice effect eliminated) to the *same* population, t would have to equal 2.01 for significance at the 5 per cent level of confidence and 2.70 for significance at the 1 per cent level of confidence. The obtained t of .81 indicates, therefore, that the difference of .85 between serial and consecutive order administrations of the Revised Stanford-Binet Intelligence Scales is statistically insignificant for our population. The correlation of .93 indicates both very close general agreement between the two arrangements and also high reliability of testing with both procedures.

Despite the very close correspondence between serial and consecutive testing indicated both by the comparison of means and by the correlation coefficient, there were some individual IQ discrepancies. These are revealed in Table 3.

TABLE 3

DISTRIBUTION OF IQ DIFFERENCES (SERIAL MINUS CONSECUTIVE) FOR 47 CHILDREN GIVEN BOTH SERIAL AND CONSECUTIVE ADMINISTRATION OF THE REVISED STANFORD-BINET SCALE

IQ Changes	Positive	Negative	Total
0 to 4	11	10	21
5 to 9	10	9	19
10 to 14	1	3	4
15 to 19	1	1	2
20 to 24	1	0	1
Total	24	23	47

The average of individual differences between IQ's obtained from serial and consecutive arrangement of the Revised Stanford-Binet Scales is 5.48, which is only very slightly higher than the average test-retest difference of 5.09 obtained by Terman and Merrill for individuals within the range of average intelligence [10, p. 46]. Eighty-five per cent of the differences, as is shown in Table 3, are within ± 10 points; approximately 15 per cent, however, are outside this range.

SUMMARY AND CONCLUSIONS

Clinicians using the Stanford-Binet Intelligence Scales extensively have claimed that serial presentation and other adaptations of prescribed Stanford-Binet procedure have many advantages over presenting the items in the consecutive order in which they appear in the manual. In the absence of experimental evidence to justify such departures from the procedures used when standardizing the scale, Terman and Merrill pointed out that such adaptations might invalidate the use of the norms for the test. Three experiments [3, 5, 7], however, now indicate that for normally

adjusted children (a) the means of IQ's obtained from serial administration do not differ significantly from those obtained from following the order prescribed in the manual; (b) between the two methods of testing, the correlation is very high, and (c) although for some children the IQ discrepancies between adaptive and consecutive testing are significant, the average IQ difference approximates the average test-retest difference for standard testing. For poorly adjusted children, however, Hutt found adaptive testing to yield higher and probably more valid IQ's.

In the light of these findings, clinicians who favor adapting the Stanford-Binet for serial testing apparently need have no hesitancy in using Terman and Merrill's norms for interpreting IQ's thus obtained.

There is, of course, one generally recognized [3] limitation, at least for the novice, in adaptive or serial testing. More on-the-spot scoring and general familiarity with the tests and their directions will be required, which means that only examiners of considerable skill should expect to employ successfully such deviations from the procedure prescribed in the manual.

Received October 31, 1949.

References

1. Bronner, Augusta F. *et al. A manual of individual mental tests and testing.* Boston: Little, Brown, 1932.
2. Harriman, P. L. Irregularity of successes on the 1937 Stanford-Binet. *J. consult. Psychol.*, 1939, **3**, 83-85.
3. Hutt, M. L. A clinical study of "consecutive" and "adaptative" testing with the Revised Stanford-Binet. *J. consult. Psychol.*, 1947, **11**, 93-103.
4. Louttit, C. M. *Clinical psychology.* New York: Harpers, 1936.
5. Louttit, C. M. *Clinical psychology.* (Rev. Ed.) New York: Harpers, 1947.
6. Mateer, Florence. *The unstable child.* New York: Appleton, 1942.
7. McCullough, Betsey R. The effects of serial testing upon the results of the Stanford-Binet Tests of Intelligence. Master of Science Thesis, Utah State Agricultural College, 1948.
8. Spache, G. Serial testing with the revised SB Scale, Form L, in the test range II-XIV. *Amer. J. Orthopsychiat.*, 1942, **12**, 81-86.
9. Terman, L. M. *The measurement of intelligence.* Boston: Houghton Mifflin, 1916.
10. Terman, L. M., and Merrill, Maude A. *Measuring intelligence.* Boston: Houghton Mifflin, 1937.
11. Wells, F. L. *Mental tests in clinical practice.* Yonkers, N. Y.: World Book Co., 1927.

THE TEACHING OF PSYCHOTHERAPY THROUGH CONTENT-FREE INTERVIEWS

CLARE WRIGHT THOMPSON
UNIVERSITY OF CALIFORNIA MEDICAL SCHOOL AND THE LANGLEY PORTER CLINIC
AND
KATHERINE BRADWAY
STANFORD UNIVERSITY

PRACTICING psychotherapists are becoming increasingly aware of the aspects of therapy which are independent of the content of the interview: the nuances of tone, the emotional interchange, the expressive movements, as opposed to the words the patient chooses to say. Teachers of psychotherapy find it difficult to offer adequate help with this kind of material. The student can prepare fairly full notes on what was said by each of the participants, but the inexperienced therapist is largely unaware of the nonverbal elements and the supervisor can only surmise what really transpired at the personal level. To stimulate the student to look for other facets of the interview, and to provide an actual learning situation to increase his skill in doing so, we have found the method presented below to be helpful.

Two students decide who is to be patient and who the therapist. They also agree upon whether the interview is initial, or whether they were previously acquainted. They then conduct an interview in which they use no words, but count, or use numbers in any sequence, instead. Each of the participants in such an interview feels quite certain that he knows what is happening in the interview, and there is a fair agreement between the two when they later discuss their impressions. In listening to such interviews, we also felt fairly certain of what was occurring, and decided to use observation as well as participation as a teaching situation.

To illustrate that this nonverbal communication does occur, and to determine to what degree the participants' impressions could also be experienced by observers, one of us carried out the following design.

A beginning therapy class of ten members, toward the end of the first quarter of instruction, played out five situations. For each situation there was one therapist, one patient, and eight observers. At the conclusion of each interchange, each student wrote his impressions according to the following instructions:

> A number of therapeutic situations are to be presented. Observe closely and after each presentation answer the following questions about what just went on:
>
> 1. What was the attitude of the patient?
> 2. Describe briefly the course of the emotional exchange between patient and therapist, paying attention to the feelings of both.
> 3. With what feelings did the patient leave?

Participation in the various interviews was by lot, with each student participating once. The participants were told:

> You are to act out this situation using numbers (counting) instead of words. It is not necessary to determine beforehand what course the interview will take. The audience has been told which of you is the patient, but not what the situation is.

The situations are given below. Three minutes were allowed for each.

> 1. This is a Child Guidance Clinic. A mother and a young child have been in treatment for one month. The mother is ready to stop treatment because the child's behavior has become worse at home.
> 2. The situation is that of a patient somewhere along in therapy asking his therapist for personal information.
> 3. A patient comes in, for the last time, to see a therapist who will no longer be able to see the patient, because the therapist is moving to another part of the country. It has been arranged, at a previous session, that the patient will be transferred to another therapist at the same clinic.

4. There is something about his past that the patient has been wanting to tell the therapist for a long time. He has determined to tell him today.

5. The patient has just been admitted after having been kept waiting 30 minutes. The previous session was broken by the therapist on short notice.

A wire recording of each interview was made. The description written by each therapist was clipped to that written by his patient, in five pairs. This description did not include the situation, only the affective interchange. The one of us unfamiliar with the students and with the details of the experiment (including what the situations were) attempted to match these descriptions with the recordings. This attempt was successful beyond the five per cent level of significance, with three interviews correctly matched and two interchanged. Another psychologist[1] with considerable therapeutic experience was asked to match the same descriptions with the recordings. His success was identical with the first matching. Neither psychologist was able to identify the first situation; once it was confused with interview four, once with interview five. That this degree of agreement should occur twice is significant beyond the one per cent point.

Samples of audience reaction ranged from, "There seemed to be little change in affect on the part of either patient or therapist. The lack of emotional tone in this interview makes it impossible for the writer to analyze the interview" to such analyses (all to situation 3) as:

At beginning, patient feels terribly hurt and lonely, and won't admit hostility. Therapist doesn't know quite what to do, and patient realizes this. Patient leaves, feeling rejected.

Patient sounds resigned and it is felt that the aura of relaxation is probably a defense for a somewhat threatening situation. The therapist has difficulty in understanding this attitude on the part of the client and as a result the free exchange of feelings is handicapped. Blocking of patient's statements results. The therapist shows some anxiety as a result of this situation. Patient maintains a consistent, non-affective emotional tone.

Patient's beginning attitude was one of indifference, arrogance, not seeming to care. Patient seemed to be indifferent, and didn't seem concerned about whether he was here or some other place. The therapist was anxious to understand the feelings of the patient, but couldn't seem to quite penetrate. I had the intuition that the therapist wanted him to come back again to see if he couldn't fare better next time.

Each of these 40 statements (made by 8 individuals to 5 situations) was typed on a separate sheet. All feminine pronouns were changed to masculine, and each paragraph was divided into three numbered parts, corresponding to the three questions asked (see above). Each of the two psychologists participating in this part of the experiment then tried (separately) to sort the statements into the five situations, matching them with the statements by the patient and therapist, and with his remembered impressions from having listened to the recordings.

These overall sortings were successful beyond the one per cent level of significance. The situations (two and three) most easily identified from the recordings were also those to which the observers' statements were most easily matched. Apparently, then, a content-free interview does give rise to certain feelings which can be communicated to a greater or a lesser degree to an audience.

Conducting an interview devoid of words seemed an enlightening experience for the students. The fact that emotional interplay without content was sufficiently important to merit an entire class session, with "patients" and "therapists" counting to each other, in itself emphasized the importance of such interplay. The students' comments about subsequent therapeutic role-taking paid greater consideration to the feelings revealed by the participants.

There was a positive relation between the instructor's ranking of the ten students, according to predicted aptitude for doing psychotherapy, and the correctness with which their descriptions of what took place during the "interviews" were matched with the statements by the patient and therapist of what took place. This suggests that the technique may have value in contributing to an appraisal of students' potential psychotherapeutic abilities. The relation was higher at the upper end than at the lower end of the scale of ranks assigned by the instructor. The students who received the three highest ranks were three of the four whose descriptions most frequently were matched correctly.

[1]We are indebted to James O. Palmer for his assistance.

It is our feeling that these content-free interviews are valuable teaching devices because they force the "overideational pre-Ph. D." (with apologies to Rapaport) to examine the intangibles which are so foreign to his academic experience. The most telling criticism of the method is that it is play-acting rather than real. To this one can say only that it has been our experience, and the experience of the unbelievers who have tried it, that when, in a content-free interview, one takes the role of therapist, he feels like a therapist. When he takes a patient role, he feels like a patient.

Received October 31, 1949.

THE MMPI AS DIAGNOSTIC DIFFERENTIATOR: A REPLY TO RUBIN

BERNARD S. AARONSON AND GEORGE S. WELSH
UNIVERSITY OF MINNESOTA

IN a previous issue of this Journal, Harold Rubin [6] published an article purporting to show that the Minnesota Multiphasic Personality Inventory was unable to differentiate among chronic alcoholics without psychosis, psychopaths, psychoneurotics, and psychotics at a veterans hospital in Ohio. In establishing these results, he took the mean scores obtained by the above groups on the subtests of the Multiphasic and submitted them to an analysis of variance testing the null hypothesis that the four clinical groups were, on any of the individual scales, representative of the same population. He rejected pattern analysis and reported that Gough's study [1] on pattern analysis in the MMPI did not hold up in his experiment.

METHOD

In an effort to demonstrate that the differentiations that Rubin denied were in fact present in his published data, two experiments were performed. In the first, the four *K*-corrected mean profiles reported by Rubin were plotted onto standard Multiphasic Profile blanks with the code of each profile according to the Welsh revision [8] of the Hathaway coding system [2] placed on each blank. Twelve clinicians were then asked to sort the profiles into the appropriate categories.

In the second experiment, the four *K*-corrected profiles, corrected to the nearest whole *T*-score, were plotted onto sheets of paper and seventeen other judges were asked to state which of the four categories the particular profile best fit. The judges were told that these were the profiles of four clinical populations, but any information as to the composition of the groups and whether they were similar or different was carefully withheld. The results were tested by Chi-square against the hypothesis that one-fourth of the designations could be correct by chance.

RESULTS

In the first group, where twelve judges sorted the profiles, there was 100 per cent success. Not a single judge made an error.

TABLE 1
CORRECT AND INCORRECT MATCHINGS OF MULTIPHASIC PROFILES COMPARED WITH CHANCE EXPECTATIONS BY CHI-SQUARE

Profile	Actual Corr.	Actual Incorr.	Chance Corr.	Chance Incorr.	χ^2	P
Psychopathic	13	4	4.25	12.75	24.0196	.01
Psychoneurotic	12	5	4.25	12.75	18.8448	.01
Psychotic	15	2	4.25	12.75	36.2573	.01
Alcoholic	10	7	4.25	12.75	10.3717	.01-.02

In the second group, where the judges had to identify the category which the profiles best fit, identifications were made correctly at far above the .01 level of confidence for the psychopathic, psychoneurotic, and psychotic profiles and between the .01 and .02 levels for the alcoholic group (Table 1). The alcoholic profile was most often misidentified as psychopathic, and the neurotic and psychopathic profiles most often as alcoholic (Table 2).

TABLE 2
IDENTIFICATIONS OF THE MEAN PROFILES OF FOUR CLINICAL GROUPS AS PSYCHOPATHIC, PSYCHONEUROTIC, PSYCHOTIC, AND ALCOHOLIC

Identifications	Profiles: Psychopathic	Psychoneurotic	Psychotic	Alcoholic
Psychopathic	13	1	1	1
Psychoneurotic		12		
Psychotic	1		15	1
Alcoholic	3	4	1	10

No judge in the second group judged incorrectly on all the profiles and almost half the

judges had scores of 100 per cent. The mean number of correct judgments was 2.94 with a SD of 1.11. When tested by the *T*-test against a hypothetical random population of judges, these results are significant at the .01 level of confidence.

DISCUSSION

The above results indicate that the differentiations which Rubin said were impossible to make were already present in his published data. Since the ratings of our judges were founded on patterns, our results also vindicate that pattern analysis which Rubin decries. They also make Rubin's attempt to use his results in an attack on conventional, so-called Aristotelian diagnosis far less convincing than he intended since his attack was based on the hypothesis that the Multiphasic does not distinguish between the diagnostic categories that he used.

Apart from whether or not his results hold, however, certain additional considerations are posed by the method which he employed. At no point does Rubin present any rationale for the choice of an analysis of variance on the individual scales of the MMPI as the best method of determining the validity of the test. It is difficult to understand why having judges match individual, mean, or median profiles with diagnoses would not have suited his purposes just as well. Certainly the results cited in the previous section of this paper would attest to the superiority of this kind of method over that which he used. However, even with his analysis of variance technique, he probably would have gotten better results had he used raw rather than *T*-scores.

In terms of the categories upon which Rubin elected to test the Multiphasic, the choice of *chronic alcoholism* is particularly unfortunate since it is a limited, phenotypic designation which does not have the same theoretical implication as the far broader classes of neurotic, psychotic, and psychopathic personality. In many modern psychiatric works, chronic alcoholism is regarded as merely a symptom of neurotic or psychopathic maladjustment [4, 5, 7]. In view of this, it is probably not altogether due to chance that the judges used in this study erred most on the alcoholic profile and tended to judge the other profiles as alcoholic when they misjudged them.

As regards the composition of his diagnostic categories, it should be noted that Rubin includes only schizophrenics in his psychotic group, while ten anxiety states, four hysterics, one hypochondriac, one neurasthenic, one reactive depressive, one mixed type, and ten unidentified individuals comprise his neurotics. The extreme heterogeneity of the one group as compared with the homogeneity of the other raises serious doubts as to the comparability of the two groups.

Finally, from the height of the mean ? scores alone, one might suspect that at least some of the profiles used were invalid. In the psychotic group, for instance, the mean *T*-score of 59.25 indicates that, on the whole, almost 12 per cent of the items were not answered. In spite of an explicit warning in the MMPI manual [3] that a score of this magnitude invalidates *any* record, Rubin explicitly states that he used seven cases in the psychotic and two in the psychoneurotic groups scoring above 70 on the ? scale. As a result, it is a tribute to the test that the judges used in this study were able to differentiate at all between the two groups.

CONCLUSIONS

1. In spite of the invalidity of some of his basic data, the differentiations which Rubin asserts the MMPI does not make are present in his published data.

2. Personality on the Minnesota Multiphasic Personality Inventory must be read in terms of the particular pattern and not by the height of any given score taken by itself.

Received October 17, 1949.

REFERENCES

1. GOUGH, H. G. Diagnostic patterns on the Minnesota Multiphasic Personality Inventory. *J. clin. Psychol.*, 1946, **2**, 23-36.
2. HATHAWAY, S. R. A coding system for MMPI profile classification. *J. consult. Psychol.*, 1947, **11**, 334-337.
3. HATHAWAY, S. R., AND MCKINLEY, J. C. *The Minnesota Multiphasic Personality Inventory.* Minneapolis: Univ. of Minnesota Press, 1943.
4. HENDERSON, D. K., AND GILLESPIE, R .D. *A textbook of psychiatry.* New York: Oxford Univ. Press, 1941.
5. NOYES, A. P. *Modern clinical psychiatry.* (2nd Ed.) Philadelphia: Saunders, 1939.
6. RUBIN, H. The Minnesota Multiphasic Personality Inventory as a diagnostic aid in a vet-

erans hospital. *J. consult. Psychol.*, 1948, **12**, 251-254.

7. STRECKER, E. A., EBAUGH, F. G., AND EWALT, J. R. *Practical clinical psychiatry* (6th Ed.) Philadelphia: Blakiston, 1947.
8. WELSH, G. S. An extension of Hathaway's MMPI coding system. *J. consult Psychol.*, 1948, **12**, 343-344.

A NOTE ON "REPLY TO RUBIN"

HAROLD RUBIN
THE OHIO STATE UNIVERSITY

A CRITICAL evaluation of this writer's recent publication on the MMPI as a diagnostic criterion [3] was made by Aaronson and Welsh [1]. Their criticisms and this writer's comments can be stated as follows:

1. *Since the ratings of our judges were founded on patterns, our results. . . . vindicate that pattern analysis which Rubin decries.* This statement would indicate that its authors have either misinterpreted or added meaning not originally intended to this writer's reference to pattern analysis. In his original article, this writer stated, "It is possible that a pattern analysis may be of some value in determining the major differences between groups, but the patterns reported by Gough do not hold up in this study". Thus, the statement that he has "decried" pattern analysis appears without foundation. A comparison of the individual profiles with those reported by Gough [2] indicates that 74 per cent of the psychoneurotic group fail to reveal the "diphasic curve with unequal peaks" and 87 per cent of the psychotic group fail to reveal the "diphasic curve with approximately co-equal peaks" reported by him. Regarding the study by Aaronson and Welsh, which investigated the relationship between mean profiles and diagnosis, it appears to this writer that the validity of pattern analyses of individual profiles cannot be inferred or supported by a study of group means.

2. *At no point does Rubin present any rationale for the choice of analysis of variance on the individual scales of the MMPI as the best method of determining the validity of the test.* The statistical technique utilized by this writer has obvious justification when the nature of the study is kept under consideration. This was a test of the null hypothesis that "on any of the individual scales, the four clinical groups were representative of the same population." Thus, the ability of the individual scales of the MMPI to differentiate between various diagnostic groups offers itself as a "natural" for the use of analysis of variance technique, a most widely accepted method of testing the significance of differences between several means.

3. *He probably would have gotten better results had he used raw rather than T-scores.* The use of *T*-scores is a function of the MMPI as constructed by Hathaway and McKinley. If the raw scores offer a more meaningful interpretation, it would appear that these should be used at all times in place of the *T*-scores required by the manual.

4. *The choice of chronic alcoholism is particularly unfortunate and it is a limited, phenotypic designation. . . . .* The writer takes no credit for this diagnostic category, but wishes to emphasize that it was in use by the VA at the time the study was conducted. He agrees that it is a "limited, phenotypic designation", but would take one further step and include the "far broader classes of neurotic, psychotic, and psychopathic personality" as equally limited, phenotypic designations.

5. *The extreme heterogeneity of the one group* (psychoneurotics) *as compared with the homogeneity of the other* (psychotics [schizophrenics]) *raises serious doubts as to the comparability of the two groups.* On the basis of behavior one group is no more heterogeneous than the other, but on the basis of "labels" the psychotic group appears more homogeneous than the psychoneurotic. However, both groups exist as such on the very basis of the homogeneity attributed to them by their psychiatric diagnoses. Thus, the argument appears to be a circular one and it is just this point of view questioned in this writer's original publication.

6. *From the height of the mean ? scores alone one might suspect that at least some of the profiles used were invalid.* Although this

criticism is a valid one, it was this writer's purpose to utilize all records as representative of situations frequently encountered in daily practice. The purposeful exclusion of some of the records would detract from the practical utility of the test and certainly limit its application. (It was believed, too, that the 12 records involved would not significantly change the results.)

This writer's general criticism of Aaronson and Welsh's "reply to Rubin" is that they have based their experiments on a misinterpretation of his original publication, which was an investigation of the individual scales of the MMPI, rather than a study of pattern analysis.

Received December 15, 1949.

References

1. Aaronson, B. S., and Welsh, G. S. The MMPI as a diagnostic differentiator: A reply to Rubin. *J. consult. Psychol.*, 1950, **14**, 324-326.
2. Gough, H. G. Diagnostic Patterns on the Minnesota Multiphasic Personality Inventory. *J. clin. Psychol.*, 1946, **2**, 23-36.
3. Rubin, H. The Minnesota Multiphasic Personality Inventory as a diagnostic aid in a veterans hospital. *J. consult. Psychol.*, 1948, **12**, 251-254.

CORRECTION OF "THE RELATIONSHIP BETWEEN VARIABILITY AND ABILITY ON THE WECHSLER-BELLEVUE"[1]

FRANCIS M. GILHOOLY
FORDHAM UNIVERSITY

DUE to an error in calculation, the figures reported for the correlation ratios describing the relationship between variability and ability on the Wechsler Bellevue Scale were incorrect. Correction of this error yields a value of .15 for eta_{yx}, showing the dependence of ability upon variability. Eta_{xy}, which gives the changes in variability resulting from changes in ability or IQ, is .25. The Pearson *r* between variability and IQ was found to be —.03. None of these values is significantly larger than zero, nor are the correlation ratios significantly different from *r*.

The results of this study thus indicate no significant relationship between variability and ability on the Bellevue Scale in a group of adult psychoneurotics. Apparently the same interpretation of scatter should be followed throughout the IQ range, i.e., the same amount of subtest scatter should be as meaningful at one level of intelligence as at another.

Received April 21, 1950.
Early publication.

[1] GILHOOLY, F. M. The relationship between variability and ability on the Wechsler-Bellevue. *J. consult. Psychol.*, 1950, **14**, 46-48.

NEW BOOKS AND TESTS

Books

ALLPORT, GORDON W. *The individual and his religion.* New York: Macmillan, 1950. Pp. xi+147. $2.50.

In one of the most thoughtful books of its field since James's classic, Professor Allport explores the origin of the quest for religion, the religions of youth and of maturity, conscience and mental health, and the natures of doubt and faith. Clinical psychologists will find the chapter on mental health interesting, but not always convincing. That "the crux of mental health. . . .is found in the nature of the individual's beliefs" checks with modern understandings. The thesis that psychotherapy has its historical roots in the work of the priest rather than of the physician is also supported intelligently. But many will disagree that "offering spiritual advice and rules of life" is a well-based mental hygiene activity. The problem of the anxiety-producing conscience is dispatched rather than clarified by calling such a conscience "immature." The plea for the coordination and cooperation of clergymen, psychologists and psychiatrists in the service of mental hygiene will, of course, receive universal support.

CHEVIGNY, HECTOR, AND BRAVERMAN, SYDELL. *The adjustment of the blind.* New Haven: Yale Univ. Press, 1950. Pp. xvi + 320. $4.00.

A very readable book that challenges the present attitudes of the general public, and social agencies in particular, to the problems of the blind. The authors have opened up the whole area of fables and false beliefs that cling to the subject of blindness. They have shown how many of these beliefs are simply the projection of the feeling about blindness by man. Society has identified understanding with sight and so questions the ability of the blind to function successfully on the levels of physical, mental, emotional and moral development. To combat these false beliefs, the authors have given from their experience the positive ways in which "re-organization" does take place. The book is based upon the assumption "that sightlessness does not create a type of response new to psychology; that the problems of personality among those who cannot see do not differ in kind from those to be encountered in the generality of human experience." It is a very stimulating book and should be thoughtfully read by social workers, psychologists, educators, and anyone who is in contact with those who are sightless or who have very limited vision.—*B.M.L.*

FREEMAN, FRANK S. *Theory and practice of psychological testing.* New York: Holt, 1950. Pp. xxi + 518. $3.50.

An introductory survey of all kinds of psychological tests: individual and group, objective, projective and situational. The coverage is encyclopedic, including descriptions of many little-used techniques as well as of the common ones. The clarity and accuracy of the book fall somewhat below its scope. Some descriptions of tests are so greatly abbreviated as to be incomprehensible to a student who does not have access to the materials. Even worse, there are a number of errors of fact. Wechsler's method of deriving IQs for the Bellevue scale, for example, is described incorrectly. As a whole, the book does not compare favorably with other texts for the first course in psychological measurement.

GUILFORD, J. P. (Ed.) *Fields of psychology.* (2nd Ed.) New York: Van Nostrand, 1950. Pp. xiv + 779. $5.00.

Fields of psychology was originally designed as a text for students in the second course in psychology. The current volume aims to bring the 1940 edition up to date. Twelve psychologists outstanding in their fields have written twenty chapters summarizing developments in such fields as animal, child, educational, abnormal, clinical, and vocational psychology, among others, and the aggregate is an excellent source book not only for students but for psychologists seeking brief orientation in fields other than their own. In a volume of this type, it was inevitable that some of the treatment would be oversimplified, while others would be highly technical; that some well-known experiments would be mentioned several times by different authors, as was the case with the Lewin, Lippitt and White study on group atmospheres; and that different authors will express contradictory points of view. The latter was particularly true with regard to psychoanalytic techniques. Almost a thousand references and literature

NOTE: Some reviews in this issue were prepared by the Associate Editors, who may be identified by their initials. Unsigned reviews are by the Editor.—*L.F.S.*

citations are listed in this volume. One wonders whether it was necessary, in a revision which purports to report the latest development in fields of psychology, to list 50 per cent of references that are 15 or more years old, many of them 25 and 30 years old, and some very much older. In spite of these weaknesses, *Fields of psychology* is an important text and reference work. —*M.K.*

KATZ, DAVID *Gestalt psychology.* (Trans. by Robert Tyson) New York: Ronald, 1950. Pp. x + 175. $3.00.

This translation from the German is not a comprehensive treatment of Gestalt psychology, but rather a series of brief essays, some almost "thumbnail" sketches, strung together into a book. The result is a very uneven performance. Some of the discussions of experimental problems are interesting, but the general treatment is overly terse, leans toward a somewhat superficial, authoritative pronouncement, and fails to place its problems in a context of current available experimental knowledge. —*W. A. H.*

LAWTON, GEORGE *How to be happy though young.* New York: Vanguard, 1949. Pp. xx + 300. $3.00.

The author organizes suggestions about the more prevalent youth problems in the form of answers to real letters written to him by boys and girls. Chapters have headings such as "Yourself," "Your Family," "People," "The Opposite Sex," "Your Career," and "The Universe." Many of the letters reproduced were written after a discussion of the problems with groups of high school students. The style is popular and the approach sound. The book should be a contribution to the young adolescent during his transition to adulthood. — *F. McK.*

LORAND, SANDOR (Ed.) *The yearbook of psychoanalysis.* New York: International Universities Press, 1949. Volume 5. Pp. 317. $7.50.

This is the fifth volume in the series, the intent of which is apparently to collect in one book some significant publications in the field during the pertinent year. The volume lacks an introduction or any specification of the criteria by which the articles were selected, and is marred by very poor proofreading. The selections are uneven in scope and content, ranging from two to thirty-two pages, from so minute a point as the meaning of losing teeth in dreams, to broad discussions of theory. Probably of greatest interest to psychologists are the symposium on the evaluation of therapeutic results, in which the participants were Oberndorf, Greenacre and Kubie, the two discussions of instinct, and a discussion of heredity and psychoanalysis, by van der Waals. The problems discussed in the symposium apply equally well to evaluation of results of psychological therapies. A point emphasized by Greenacre is relevant to discussions of research versus practice: "It seems to be so simply true as to be elusive, that the worker whose goal is the essential verity of his scientific work may in some instances take unnecessary detours of exploration, but by and large will contribute most to the science and to his patients." Alexander is represented by a critical discussion of the theory of life and death instincts, in which he interprets what Freud called the death instinct as "disintegration of mature behavior into its elementary parts." Kubie's paper on instincts and homeostasis, reprinted from *Psychosomatic Medicine,* is an attempt at classification of instincts based on experimental biology and psychology rather than on analytical theory, and is a contribution of major importance in this area. —*A. R.*

MAY, ROLLO *The meaning of anxiety.* New York: Ronald Press, 1950. Pp. xv + 376. $4.50.

With anxiety so central a concept in our understanding of the psychoneuroses, it is extraordinary that so few books have dealt with it comprehensively: Kierkegaard's in 1844, Freud's in 1926, and May's in 1950. The core of the present volume is a series of chapters that summarize and interpret views of anxiety as seen philosophically, biologically, psychologically, and culturally. The philosophical background is an especially useful section, for the material is rarely familiar to psychologists, except for the few also trained in philosophy. In the chapter on psychological interpretations of anxiety, the experimental studies from Watson to Mowrer are considered first, critically but not unsympathetically. The analytic theories from Freud to Sullivan follow, with the author leaning toward the neo-analysts. In his summary and synthesis, May attempts his own definition of anxiety: "the apprehension cued off by a threat to some value which the individual holds essential to his existence as a personality." Many readers will find this definition somewhat less than satisfying, especially because many of its essential component words—"apprehension," "threat," "value," "existence"—are not themselves defined strictly. Part II of the book consists of fourteen case studies demonstrating anxiety. One case is a well-educated young man; the remaining thirteen are unmarried pregnant young women awaiting the birth of their babies. Although the study of the latter group is almost a research, the cases serve more to illustrate the author's theories than to provide a critical test of them. As instructional material, May's book has fine values. It is graciously written, scholarly and temperate in tone, and widely informative. If it does not settle all the problems of anxiety, the fault lies more with our science than with the author. The book should be read by all clinical psychologists and by their students.

SENN, MILTON J. E. (Ed.) *Problems of infancy and childhood.* Transactions of the Third Conference. New York: Josiah Macy, Jr. Foundation, 1949. Pp. 156.

A report of a conference of twenty-five outstanding psychiatrists and pediatricians. The discussion centered about two research reports, one presented by Dr. Sibylle Escalona of the Menninger Foundation, on the psychological situation of mother and

child upon return from the hospital, and a second by Drs. Lucie Jessner and Samuel Kaplan of the Psychiatric Service for Children of the Massachusetts General Hospital, on the emotional reactions of children to tonsillectomy and adenoidectomy. Although the research reports present new and important data, the publication is of particular interest because of the detailed recording of the discussions of the highly qualified participants. — *M. K.*

SPEARMAN, C. AND JONES, LL. WYNN *Human ability.* London and New York: Macmillan, 1950. Pp. vii + 198. $2.50.

Professor Spearman's final formulation of his theory of abilities was planned as a continuation of his *Abilities of man* (1927). It reveals a narrowing of the gap between his original two-factor theory and the multi-factor theories popular in America, especially in the discussions of the "broad" specific factors, including the verbal and the mechanical. The nonintellective factors such as perseveration, fluency and oscillation receive brief mention, reminding us again of challenging areas of human behavior that have yet to be explored adequately.

STONE, CALVIN P. (Ed.) *Annual review of psychology.* Stanford, Calif.: Annual Reviews, Inc., 1950. Vol. I. Pp. x + 330. $6.00.

The *Annual reviews* promise to bridge the gap between a periodical literature that has grown too vast to be read fully, and textbooks that often lag years behind the frontiers of research. Volume I covers 18 areas of psychology by chapters that average 17 pages and that cite a total of 1594 references. The unity of the science of psychology stands out strikingly in the volume, for there is hardly a chapter without relevance to clinical concerns. The sections on growth, development and decline (Jones and Bayley), on learning (Melton), on individual differences (Thorndike), and on social and group processes (Bruner) are no less important for well-founded clinical theory than are the chapters which deal more overtly with applied material. Chapters differ considerably in their authors' conceptions of objectives and scope. Extremes in this respect are illustrated by Sears's sharply and splendidly critical chapter on personality with only 39 references, and Cameron's useful but more enumerative one on abnormalities of behavior which cites 142 sources. For reasons that seem quite artificial, the clinical-counseling field is broken in two, with two chapters on *clinical methods,* psychodiagnostics (H. F. Hunt) and psychotherapy (Snyder), and two on *counseling methods,* diagnostics (Berdie) and therapy (Borden). The resulting duplications document quite clearly the unity of "clinical" and "counseling" psychology, and it is hoped that future volumes will not perpetuate the unnecessary separation for too many years. There are a few errors as well as repetitions. For example, W. A. Hunt's studies of abbreviated tests are described correctly by H. F. Hunt on page 215, and somewhat incorrectly by Berdie on page 257. Repetitions and mistakes may perhaps be excused by the inexperience of the editors and by the time pressures under which they worked, for which they apologize in the Preface. Despite a few shortcomings, however, the *Annual review* is a notable achievement, deserving of careful reading both by graduate students and by experienced psychologists. The series will become a *must* on our reference shelves.

TAFT, DONALD R. *Criminology.* (Rev. Ed.) New York: Macmillan, 1950. Pp. xiv + 704. $5.50.

An elementary textbook on the problems of crime and criminals that is notable for its almost complete ignoring of the psychological approach. The descriptive and historical material on crime in various groups, and on prisons and other punishments, is presented interestingly and will be informative to most psychologists.

Fostering mental health in our schools. Association for Supervision and Curriculum Development, 1950 Yearbook. Washington: National Education Association, 1950. Pp. ix + 320. $3.00.

A committee of educators has prepared an attractive and useful volume that tries to outline a program for the promotion of good mental health, in distinction to the treatment of psychological deviations. While lacking in basic theory, the program recommends much that is sound and little with which any psychologist can take serious issue. Practical workers will be helped by the many illustrative anecdotes that make the meanings clear even to those with little formal preparation in psychology.

TESTS

BLUM, GERALD S. *The Blacky Pictures.* Ages 5-adult. Individual test. Untimed (45) min. 12 plates, 2 books of inquiry cards, manual, pp. 24 ($9.00); record blank ($3.50 per 25). New York: Psychological Corp., 1950.

Blum's new projective test revolves around the psychoanalytically-oriented adventures of Blacky, a small dog whose sex is identified as that of the examinee. The cartoon cards are structured to suggest specific problems of psychosexual development, including oral eroticism, oral sadism, anal sadism, sibling rivalry, guilt feelings, and others. A free narrative is followed by an inquiry with the use of multiple-choice questions for each picture, printed on cards. Although the amusing cartoons suggest a children's test, the technique was first used in a research with college students, and has been administered to about 400 adult subjects. Preliminary trials suggest that it can evoke responses from children as young as five, and the manual provides a suggested alternative inquiry for children. Although the manual and the research monograph (*Genet. Psychol. Monogr.,* 1949, **39,** 3-99) suggest the pur-

poses and uses of the test, no attempt is made to provide full training for its interpretation. The author wisely suggests that it be used only by clinical psychologists with general competence in other projective tests and with a broad knowledge of psychoanalytic theory.

DELP, H. A. *The Delp IQ Computer.* Boston: Houghton Mifflin, 1950. $2.00.

A circular slide rule of plastic material, 6 inches in diameter, to aid the calculation of CAs and IQs. One side is used to calculate months of CA since the last birthday, from the birth date and the testing date. The other side of the computer divides MA by CA to obtain IQ, with a special scale for CAs between 13 and 16 for use with the Revised Stanford-Binet.

FREEMAN, ELLIS *The Illuminant-Stable Color Vision Test.* Child-adult. Individual test. Book of 12 plates, with manual, pp. 14 ($15). New York: Psychological Corp., 1949.

A major difficulty encountered with all previous color blindness tests employing printed plates has been the varying results obtained under different types of illumination. The "Illuminant-Stable" test uses inks that maintain the same relationships under a wide variety of lights, from yellow incandescents to bright daylight. The first six plates constitute the main test, and separate all color-deficients from normals. The normal eye sees a number in a colored geometric confusion pattern that the color-blind eye cannot see. Six additional plates are provided for experimental use in separating protanopes from deuteranopes.

ROTTER, JULIAN B. *The Rotter Incomplete Sentences Blank.* College students. 1 form. Untimed (20) min. Test blank ($1.10 per 25); manual by J. B. Rotter and Janet E. Rafferty, pp. 86, ($1.70). New York: Psychological Corp., 1950.

The Rotter blank consists of 40 incomplete sentence items, mostly with very short stems (I like . . . .; My mind. . . .; Most girls. . . .), that have shown much promise as a method of personality study. Scoring is made relatively objective by the use of numerous scoring examples, which are separately tabulated for men and women. The manual contains six practice records for training in scoring, and six cases which are discussed interpretively. The reliability of the test is .84, and its validity for separating groups of well-adjusted and maladjusted students is indicated by biserial correlations of .50 for 82 women and .62 for 124 men. Blanks, with a few items revised appropriately, are available for high school and adult levels, but a manual has been prepared only for the college group.

TIEGS, ERNEST W., AND CLARK, WILLIS W. *California Reading Test.* Intermediate, grades 7-9. 4 forms. 50 (58) min. IBM, hand, or automatic scoring. Test booklet ($1.50 per 25), with manual, pp. 16; Scoreze answer sheets for IBM and automatic scoring ($1.75 per 25). Los Angeles: California Test Bureau, 1950.

The *California Reading Test,* formerly the *Progressive Reading Test,* may be used alone or as part of a battery with corresponding arithmetic and language sections. The reading test consists of two main parts: vocabulary, with items from mathematics, science, social science, and general vocabularies; and comprehension, including following directions, reference skills, and paragraph interpretation. Reported reliabilities in a single grade group are high for each major part and for the whole: .90, .89, and .92, respectively. Validity is defended in terms of the processes used in the selection of items. Percentile norms are given for half-grade groups, and scores may also be interpreted in terms of reading-grade and reading-age norms. The test and manual formats, redesigned by Raymond Loewy Associates, are attractive but violate at least a few experimentally determined findings about legibility, such as the use of white letters on a colored background. The new "Scoreze" answer sheets combine an IBM machine-scoring form for supervisory use with a carbon paper self-scoring device for immediate interpretation by the teacher.

Recently Published

PSYCHOLOGICAL MONOGRAPHS: GENERAL AND APPLIED

Volume 63, 1949

Facial Expressions of Emotion. James C. Coleman, University of Southern California. #296, $1.00

A Comparative Study of the Wherry-Doolittle and a Multiple Cutting-Score Method. Glen Grimsley, General Motors Institute. #297, $.75

Factor Analyses of Tests and Criteria: A Comparative Study of Two AAF Pilot Populations. William B. Michael, Princeton University. #298, $1.00

The Appraisal of Parent Behavior. Alfred L. Baldwin, Joan Kalhorn, and Fay Huffman Breese, Fels Research Institute. #299, $1.50

Studies of Identical Twins Reared Apart. The late Barbara S. Burks; and Anne Roe, New York City. #300, $1.00

Color Preferences of Psychiatric Groups. Samuel J. Warner, New York City. #301, $.75

Perception of Body Position and of the Position of the Visual Field. H. A. Witkin, Brooklyn College. #302, $1.00

An Experimental Examination of the Thematic Apperception Technique in Clinical Diagnosis. A. A. Hartman, Boston University. #303, $1.00

Religion and Humanitarianism: A Study of Institutional Implications. Clifford Kirkpatrick, Indiana University. #304, $.75

The Development and Validation of a Set of Musical Ability Tests. Robert W. Lundin, Hamilton College. #305, $1.00

A Comparison of Two Tests of Intelligence Administered to Adults. Anna S. Elonen, University of Chicago. #306, $1.00

The 1949 volume of the Psychological Monographs consists of eleven separate issues. Orders for any of these can be placed separately at the prices listed above, or the entire volume can be ordered for $6.00.

American Psychological Association

1515 Massachusetts Avenue N.W., Washington 5, D.C.

PSYCHOLOGICAL MONOGRAPHS: GENERAL AND APPLIED

Volume 64, 1950

Patterns of Personality Rigidity and Some of Their Determinants. Seymour Fisher, Elgin State Hospital. ‡307, $1.00

The Value of an Oral Reading Test for Diagnosis of the Reading Difficulties of College Freshmen of Low Academic Performance. Charles A. Wells, American International College. ‡ 308, $1.00

Rorschach Responses Related to Vocational Interests and Job Satisfaction. Solis L. Kates, Michigan State College. ‡309, $1.00

Symbol Elaboration Test (S.E.T.): The Reliability and Validity of a New Projective Technique. Johanna Krout, Chicago Psychological Institute. ‡310, $2.00

Changes in Responses to the Minnesota Multiphasic Inventory Following Certain Therapies. William Schofield, University of Minnesota. ‡311, $1.00

A Scale for Measuring Teacher-Pupil Attitudes and Teacher-Pupil Rapport. Carroll H. Leeds, Furman University. ‡312, $1.00

The Nature and Efficacy of Methods of Attack on Reasoning Problems. Benjamin Burack, Roosevelt College. ‡313, $1.00

The Validity of a Multiple-Choice Projective Test in Psychopathological Screening. Martin Singer, VA Hospital, Long Island. ‡314, $1.00

A Normative Study of the Thematic Apperception Test. Leonard D. Eron, Yale University. ‡315, $1.50

Experimentally Induced Variations in Rorschach Performance. Edith E. Lord, Arizona State Department of Health. ‡316, $1.00

An Evaluation of Personality-Trait Ratings Obtained by Unstructured Assessment Interviews. Ernest C. Tupes, United States Air Force. ‡317, $1.00

The 1950 volume of the Psychological Monographs will consist of eleven separate issues. Orders can be placed for any of these; the orders will be filled when the issues appear. On July 6, 1950, ‡307 and ‡308 had been published.

The entire 1950 volume may be subscribed to for $6.00; it will probably be completed by December of 1950.

American Psychological Association

1515 Massachusetts Avenue N.W., Washington 5, D.C.